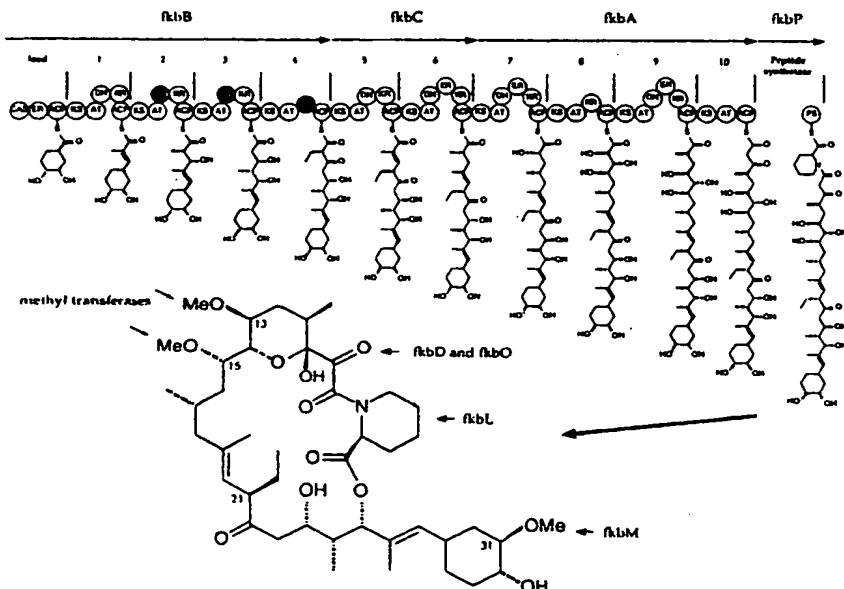




INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/52, 15/54, 15/62, 9/10, C12P 17/18, 19/32, C07D 498/18 // (C07D 498/18, 311:00, 273:00, 211:00)		A2	(11) International Publication Number: WO 00/20601 (43) International Publication Date: 13 April 2000 (13.04.00)
(21) International Application Number: PCT/US99/22886 (22) International Filing Date: 1 October 1999 (01.10.99)		(74) Agents: FAVORITO, Carolyn et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).	
(30) Priority Data: 60/102,748 2 October 1998 (02.10.98) US 60/123,810 11 March 1999 (11.03.99) US 60/139,650 17 June 1999 (17.06.99) US		(81) Designated States: AL, AM, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, IL, IS, JP, KG, KP, KR, L.C, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SI, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SI), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).		Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(72) Inventors; and (75) Inventors/Applicants (for US only): REEVES, Christopher [US/US]; 4 East Altarinda Drive, Orinda, CA 94563 (US). CHU, Daniel [US/US]; 3767 Benton Street, Santa Clara, CA 95051 (US). KHOSLA, Chaitan [IN/US]; 740 Para Avenue, Palo Alto, CA 94306 (US). SANTI, Daniel [US/US]; 211 Belgrave Avenue, San Francisco, CA 94117 (US). WU, Kai [CN/US]; 900 Constitution Drive, Foster City, CA 94404 (US).			

(54) Title: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS THEREFOR



(57) Abstract

Host cells comprising recombinant vectors encoding the FK-520 polyketide synthase and FK-520 modification enzymes can be used to produce the FK-520 polyketide. Recombinant DNA constructs comprising one or more FK-520 polyketide synthase domains, modules, open reading frames, and variants thereof can be used to produce recombinant polyketide synthases and a variety of different polyketides with application as pharmaceutical and veterinary products.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SV	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
CONSTRUCTS THEREFOR

5

Field of the Invention

The present invention relates to polyketides and the polyketide synthase (PKS) enzymes that produce them. The invention also relates generally to genes encoding PKS enzymes and to recombinant host cells containing such genes and in which expression of such genes leads to the production of polyketides. The present invention also relates to 10 compounds useful as medicaments having immunosuppressive and/or neurotrophic activity. Thus, the invention relates to the fields of chemistry, molecular biology, and agricultural, medical, and veterinary technology.

Background of the Invention

15 Polyketides are a class of compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes.

Polyketides are biologically active molecules with a wide variety of structures, and the class encompasses numerous compounds with diverse activities. Tetracycline,

20 erythromycin, epothilone, FK-506, FK-520, narbomycin, picromycin, rapamycin, spinocyn, and tylosin are examples of polyketides. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds.

25 This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that

otherwise do not produce the polyketide. The technology also allows one to produce 30 molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters. See, e.g., PCT publication Nos. WO 93/13663;

95/08548; 96/40968; 97/02358; 98/27203; and 98/49315; United States Patent Nos.

4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; 5,712,146; 5,830,750; and

5,843,718; and Fu *et al.*, 1994, *Biochemistry* 33: 9321-9326; McDaniel *et al.*, 1993,

35 *Science* 262: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by PKS enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between acylthioester building blocks. The building blocks used to form complex polyketides are typically acylthioesters, such as acetyl, butyryl, propionyl, malonyl, hydroxymalonyl, methylmalonyl, and ethylmalonyl CoA. Other building blocks include amino acid like acylthioesters. PKS enzymes that incorporate such building blocks include an activity that functions as an amino acid ligase (an AMP ligase) or as a non-ribosomal peptide synthetase (NRPS). Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesized. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

In the Type I or modular PKS enzyme group, a set of separate catalytic active sites (each active site is termed a "domain", and a set thereof is termed a "module") exists for each cycle of carbon chain elongation and modification in the polyketide synthesis pathway. The typical modular PKS is composed of several large polypeptides, which can be segregated from amino to carboxy termini into a loading module, multiple extender modules, and a releasing (or thioesterase) domain. The PKS enzyme known as 6-deoxyerythronolide B synthase (DEBS) is a Type I PKS. In DEBS, there is a loading module, six extender modules, and a thioesterase (TE) domain. The loading module, six extender modules, and TE of DEBS are present on three separate proteins (designated DEBS-1, DEBS-2, and DEBS-3, with two extender modules per protein). Each of the DEBS polypeptides is encoded by a separate open reading frame (ORF) or gene; these genes are known as *eryA1*, *eryAII*, and *eryAIII*. See Caffrey *et al.*, 1992, *FEBS Letters* 304: 205, and U.S. Patent No. 5,824,513, each of which is incorporated herein by reference.

Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The loading module of DEBS consists of an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain. Another type of loading module utilizes an inactivated ketosynthase (KS) domain and AT and ACP domains. This inactivated KS is in some instances called KS^Q, where the superscript letter is the abbreviation for the amino acid, glutamine, that is present instead of the active site cysteine required for ketosynthase activity. In other PKS enzymes, including the FK-506 PKS, the loading module

incorporates an unusual starter unit and is composed of a CoA ligase like activity domain. In any event, the loading module recognizes a particular acyl-CoA (usually acetyl or propionyl but sometimes butyryl or other acyl-CoA) and transfers it as a thiol ester to the ACP of the loading module.

5 The AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, i.e., methylmalonyl, ethylmalonyl, and 2-hydroxymalonyl) and transfers it to the ACP of that extender module to form a thioester. Each extender module is responsible for accepting a compound from a prior module, binding a building block, attaching the building block to the compound from the prior 10 module, optionally performing one or more additional functions, and transferring the resulting compound to the next module.

15 Each extender module of a modular PKS contains a KS, AT, ACP, and zero, one, two, or three domains that modify the beta-carbon of the growing polyketide chain. A typical (non-loading) minimal Type I PKS extender module is exemplified by extender module three of DEBS, which contains a KS domain, an AT domain, and an ACP domain. These three domains are sufficient to activate a 2-carbon extender unit and attach it to the growing polyketide molecule. The next extender module, in turn, is responsible for attaching the next building block and transferring the growing compound to the next extender module until synthesis is complete.

20 Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module is transferred to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module one possesses an acyl-KS and a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading building block (elongation or extension).

25 The polyketide chain, growing by two carbons each extender module, is sequentially passed as covalently bound thiol esters from extender module to extender module, in an assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the name polyketide arises. Most commonly, however, additional enzymatic activities modify the beta keto group of each two carbon unit just after it has been added to the growing polyketide chain but before it is transferred to the next module.

Thus, in addition to the minimal module containing KS, AT, and ACP domains necessary to form the carbon-carbon bond, and as noted above, other domains that modify the beta-carbonyl moiety can be present. Thus, modules may contain a ketoreductase (KR) domain that reduces the keto group to an alcohol. Modules may also 5 contain a KR domain plus a dehydratase (DH) domain that dehydrates the alcohol to a double bond. Modules may also contain a KR domain, a DH domain, and an enoylreductase (ER) domain that converts the double bond product to a saturated single bond using the beta carbon as a methylene function. An extender module can also contain other enzymatic activities, such as, for example, a methylase or dimethylase 10 activity.

After traversing the final extender module, the polyketide encounters a releasing domain that cleaves the polyketide from the PKS and typically cyclizes the polyketide. For example, final synthesis of 6-dEB is regulated by a TE domain located at the end of extender module six. In the synthesis of 6-dEB, the TE domain catalyzes cyclization of 15 the macrolide ring by formation of an ester linkage. In FK-506, FK-520, rapamycin, and similar polyketides, the TE activity is replaced by a RapP (for rapamycin) or RapP like activity that makes a linkage incorporating a pipecolate acid residue. The enzymatic activity that catalyzes this incorporation for the rapamycin enzyme is known as RapP, encoded by the *rapP* gene. The polyketide can be modified further by tailoring enzymes; 20 these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule. For example, 6-dEB is hydroxylated at C-6 and C-12 and glycosylated at C-3 and C-5 in the synthesis of erythromycin A.

In Type I PKS polypeptides, the order of catalytic domains is conserved. When 25 all beta-keto processing domains are present in a module, the order of domains in that module from N-to-C-terminus is always KS, AT, DH, ER, KR, and ACP. Some or all of the beta-keto processing domains may be missing in particular modules, but the order of the domains present in a module remains the same. The order of domains within modules is believed to be important for proper folding of the PKS polypeptides into an active 30 complex. Importantly, there is considerable flexibility in PKS enzymes, which allows for the genetic engineering of novel catalytic complexes. The engineering of these enzymes is achieved by modifying, adding, or deleting domains, or replacing them with those taken from other Type I PKS enzymes. It is also achieved by deleting, replacing, or adding entire modules with those taken from other sources. A genetically engineered

PKS complex should of course have the ability to catalyze the synthesis of the product predicted from the genetic alterations made.

Alignments of the many available amino acid sequences for Type I PKS enzymes has approximately defined the boundaries of the various catalytic domains. Sequence 5 alignments also have revealed linker regions between the catalytic domains and at the N- and C-termini of individual polypeptides. The sequences of these linker regions are less well conserved than are those for the catalytic domains, which is in part how linker regions are identified. Linker regions can be important for proper association between domains and between the individual polypeptides that comprise the PKS complex. One 10 can thus view the linkers and domains together as creating a scaffold on which the domains and modules are positioned in the correct orientation to be active. This organization and positioning, if retained, permits PKS domains of different or identical substrate specificities to be substituted (usually at the DNA level) between PKS enzymes by various available methodologies. In selecting the boundaries of, for example, an AT 15 replacement, one can thus make the replacement so as to retain the linkers of the recipient PKS or to replace them with the linkers of the donor PKS AT domain, or, preferably, make both constructs to ensure that the correct linker regions between the KS and AT domains have been included in at least one of the engineered enzymes. Thus, there is considerable flexibility in the design of new PKS enzymes with the result that 20 known polyketides can be produced more effectively, and novel polyketides useful as pharmaceuticals or for other purposes can be made.

By appropriate application of recombinant DNA technology, a wide variety of polyketides can be prepared in a variety of different host cells provided one has access to nucleic acid compounds that encode PKS proteins and polyketide modification enzymes. 25 The present invention helps meet the need for such nucleic acid compounds by providing recombinant vectors that encode the FK-520 PKS enzyme and various FK-520 modification enzymes. Moreover, while the FK-506 and FK-520 polyketides have many useful activities, there remains a need for compounds with similar useful activities but with better pharmacokinetic profile and metabolism and fewer side-effects. The present 30 invention helps meet the need for such compounds as well.

Summary of the Invention

In one embodiment, the present invention provides recombinant DNA vectors that encode all or part of the FK-520 PKS enzyme. Illustrative vectors of the invention 35 include cosmid pKOS034-120, pKOS034-124, pKOS065-C31, pKOS065-C3,

pKOS065-M27, and pKOS065-M21. The invention also provides nucleic acid compounds that encode the various domains of the FK-520 PKS, i.e., the KS, AT, ACP, KR, DH, and ER domains. These compounds can be readily used, alone or in combination with nucleic acids encoding other FK-520 or non-FK-520 PKS domains, as 5 intermediates in the construction of recombinant vectors that encode all or part of PKS enzymes that make novel polyketides.

The invention also provides isolated nucleic acids that encode all or part of one or more modules of the FK-520 PKS, each module comprising a ketosynthase activity, an acyl transferase activity, and an acyl carrier protein activity. The invention provides an 10 isolated nucleic acid that encodes one or more open reading frames of FK-520 PKS genes, said open reading frames comprising coding sequences for a CoA ligase activity, an NRPS activity, or two or more extender modules. The invention also provides recombinant expression vectors containing these nucleic acids.

In another embodiment, the invention provides isolated nucleic acids that encode 15 all or a part of a PKS that contains at least one module in which at least one of the domains in the module is a domain from a non-FK-520 PKS and at least one domain is from the FK-520 PKS. The non-FK-520 PKS domain or module originates from the rapamycin PKS, the FK-506 PKS, DEBS, or another PKS. The invention also provides recombinant expression vectors containing these nucleic acids.

20 In another embodiment, the invention provides a method of preparing a polyketide, said method comprising transforming a host cell with a recombinant DNA vector that encodes at least one module of a PKS, said module comprising at least one FK-520 PKS domain, and culturing said host cell under conditions such that said PKS is produced and catalyzes synthesis of said polyketide. In one aspect, the method is practiced with a *Streptomyces* host cell. In another aspect, the polyketide produced is 25 FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-506 or rapamycin.

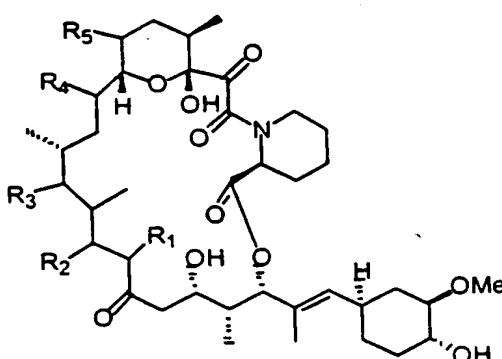
In another embodiment, the invention provides a set of genes in recombinant 30 form sufficient for the synthesis of ethylmalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require ethylmalonyl CoA for biosynthesis. The invention also provides recombinant nucleic acids that encode AT domains specific for ethylmalonyl CoA. Thus, the compounds of the invention can be

used to produce polyketides requiring ethylmalonyl CoA in host cells that otherwise are unable to produce such polyketides.

In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require 2-hydroxymalonyl CoA for biosynthesis. The invention also provides recombinant nucleic acids that encode AT domains specific for 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA. Thus, the compounds of the invention can be used to produce polyketides requiring 2-hydroxymalonyl CoA or 2-methoxymalonyl CoA in host cells that are otherwise unable to produce such polyketides.

In another embodiment, the invention provides a compound related in structure to FK-520 or FK-506 that is useful in the treatment of a medical condition. These compounds include compounds in which the C-13 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. Such compounds are less susceptible to the main *in vivo* pathway of degradation for FK-520 and FK-506 and related compounds and thus exhibit an improved pharmacokinetic profile. The compounds of the invention also include compounds in which the C-15 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. The compounds of the invention also include the above compounds further modified by chemical methodology to produce derivatives such as, but not limited to, the C-18 hydroxyl derivatives, which have potent neurotrophin but not immunosuppression activities.

Thus, the invention provides polyketides having the structure:



wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen

or hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506. The invention provides these compounds in purified form and in pharmaceutical compositions.

5 In another embodiment, the invention provides a method for treating a medical condition by administering a pharmaceutically efficacious dose of a compound of the invention. The compounds of the invention may be administered to achieve immunosuppression or to stimulate nerve growth and regeneration.

10 These and other embodiments and aspects of the invention will be more fully understood after consideration of the attached Drawings and their brief description below, together with the detailed description, examples, and claims that follow.

Brief Description of the Drawings

15 Figure 1 shows a diagram of the FK-520 biosynthetic gene cluster. The top line provides a scale in kilobase pairs (kb). The second line shows a restriction map with selected restriction enzyme recognition sequences indicated. K is *Kpn*I; X is *Xho*I, S is *Sac*I; P is *Pst*I; and E is *Eco*RI. The third line indicates the position of FK-520 PKS and related genes. Genes are abbreviated with a one letter designation, i.e., C is *fkbC*. Immediately under the third line are numbered segments showing where the loading 20 module (L) and ten different extender modules (numbered 1 - 10) are encoded on the various genes shown. At the bottom of the Figure, the DNA inserts of various cosmids of the invention (i.e., 34-124 is cosmid pKOS034-124) are shown in alignment with the FK-520 biosynthetic gene cluster.

25 Figure 2 shows the loading module (load), the ten extender modules, and the peptide synthetase domain of the FK-520 PKS, together with, on the top line, the genes that encode the various domains and modules. Also shown are the various intermediates in FK-520 biosynthesis, as well as the structure of FK-520, with carbons 13, 15, 21, and 31 numbered. The various domains of each module and subdomains of the loading 30 module are also shown. The darkened circles showing the DH domains in modules 2, 3, and 4 indicate that the dehydratase domain is not functional as a dehydratase; this domain may affect the stereochemistry at the corresponding position in the polyketide. The substituents on the FK-520 structure that result from the action of non-PKS enzymes are also indicated by arrows, together with the types of enzymes or the genes that code for the enzymes that mediate the action. Although the methyltransferase is shown acting 35 at the C-13 and C-15 hydroxyl groups after release of the polyketide from the PKS, the

methyltransferase may act on the 2-hydroxymalonyl substrate prior to or contemporaneously with its incorporation during polyketide synthesis.

Figure 3 shows a close-up view of the left end of the FK-520 gene cluster, which contains at least ten additional genes. The ethyl side chain on carbon 21 of FK-520 (Figure 2) is derived from an ethylmalonyl CoA extender unit that is incorporated by an ethylmalonyl specific AT domain in extender module 4 of the PKS. At least four of the genes in this region code for enzymes involved in ethylmalonyl biosynthesis. The polyhydroxybutyrate depolymerase is involved in maintaining hydroxybutyryl-CoA pools during FK-520 production. Polyhydroxybutyrate accumulates during vegetative growth and disappears during stationary phase in other *Streptomyces* (Ranade and Vining, 1993, *Can. J. Microbiol.* 39:377). Open reading frames with unknown function are indicated with a question mark.

Figure 4 shows a biosynthetic pathway for the biosynthesis of ethylmalonyl CoA from acetoacetyl CoA consistent with the function assigned to four of the genes in the FK-520 gene cluster shown in Figure 3.

Figure 5 shows a close-up view of the right-end of the FK-520 PKS gene cluster (and of the sequences on cosmid pKOS065-C31). The genes shown include *fkbD*, *fkbM* (a methyl transferase that methylates the hydroxyl group on C-31 of FK-520), *fkbN* (a homolog of a gene described as a regulator of cholesterol oxidase and that is believed to be a transcriptional activator), *fkbQ* (a type II thioesterase, which can increase polyketide production levels), and *fkbS* (a crotonyl-CoA reductase involved in the biosynthesis of ethylmalonyl CoA).

Figure 6 shows the proposed degradative pathway for tacrolimus (FK-506) metabolism.

Figure 7 shows a schematic process for the construction of recombinant PKS genes of the invention that encode PKS enzymes that produce 13-desmethoxy FK-506 and FK-520 polyketides of the invention, as described in Example 4, below.

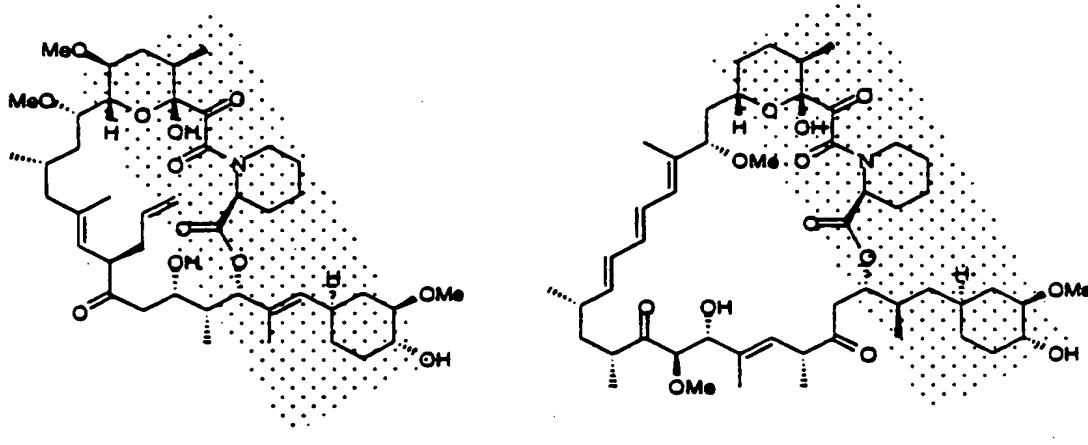
Figure 8, in Parts A and B, shows certain compounds of the invention preferred for dermal application in Part A and a synthetic route for making those compounds in Part B.

Detailed Description of the Invention

Given the valuable pharmaceutical properties of polyketides, there is a need for methods and reagents for producing large quantities of polyketides, as well as for producing related compounds not found in nature. The present invention provides such

methods and reagents, with particular application to methods and reagents for producing the polyketides known as FK-520, also known as ascomycin or L-683,590 (see Holt *et al.*, 1993, *JACS* 115:9925), and FK-506, also known as tacrolimus. Tacrolimus is a macrolide immunosuppressant used to prevent or treat rejection of transplanted heart, 5 kidney, liver, lung, pancreas, and small bowel allografts. The drug is also useful for the prevention and treatment of graft-versus-host disease in patients receiving bone marrow transplants, and for the treatment of severe, refractory uveitis. There have been additional reports of the unapproved use of tacrolimus for other conditions, including alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple 10 sclerosis, primary biliary cirrhosis, and scleroderma. The invention provides methods and reagents for making novel polyketides related in structure to FK-520 and FK-506, and structurally related polyketides such as rapamycin.

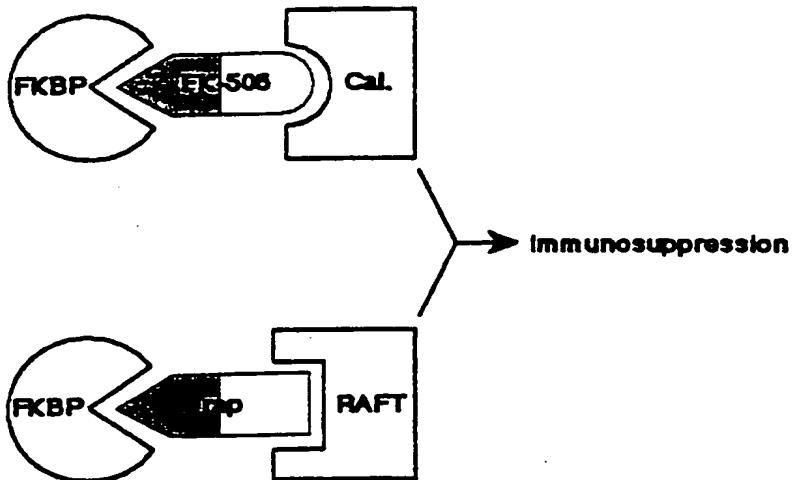
The FK-506 and rapamycin polyketides are potent immunosuppressants, with chemical structures shown below.



15 FK-520 differs from FK-506 in that it lacks the allyl group at C-21 of FK-506, having instead an ethyl group at that position, and has similar activity to FK-506, albeit reduced immunosuppressive activity.

These compounds act through initial formation of an intermediate complex with 20 protein "imunophilins" known as FKBP (FK-506 binding proteins), including FKBP-12. Immunophilins are a class of cytosolic proteins that form complexes with molecules such as FK-506, FK-520, and rapamycin that in turn serve as ligands for other cellular targets involved in signal transduction. Binding of FK-506, FK-520, and rapamycin to FKBP occurs through the structurally similar segments of the polyketide molecules, 25 known as the "FKBP-binding domain" (as generally but not precisely indicated by the

stippled regions in the structures above). The FK-506-FKBP complex then binds calcineurin, while the rapamycin-FKBP complex binds to a protein known as RAFT-1. Binding of the FKBP-polyketide complex to these second proteins occurs through the dissimilar regions of the drugs known as the "effector" domains.



5

The three component FKBP-polyketide-effector complex is required for signal transduction and subsequent immunosuppressive activity of FK-506, FK-520, and rapamycin. Modifications in the effector domains of FK-506, FK-520, and rapamycin that destroy binding to the effector proteins (calcineurin or RAFT) lead to loss of

10 immunosuppressive activity, even though FKBP binding is unaffected. Further, such analogs antagonize the immunosuppressive effects of the parent polyketides, because they compete for FKBP. Such non-immunosuppressive analogs also show reduced toxicity (see Dumont *et al.*, 1992, *Journal of Experimental Medicine* 176, 751-760), indicating that much of the toxicity of these drugs is not linked to FKBP binding.

15 In addition to immunosuppressive activity, FK-520, FK-506, and rapamycin have neurotrophic activity. In the central nervous system and in peripheral nerves, immunophilins are referred to as "neuroimmunophilins". The neuroimmunophilin FKBP is markedly enriched in the central nervous system and in peripheral nerves. Molecules that bind to the neuroimmunophilin FKBP, such as FK-506 and FK-520, have the 20 remarkable effect of stimulating nerve growth. *In vitro*, they act as neurotrophins, i.e., they promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. See Gold *et al.*, Jun. 1999, *J. Pharm. Exp. Ther.* 289(3): 1202-1210; Lyons *et al.*, 1994, *Proc. National 25 Academy of Science* 91: 3191-3195; Gold *et al.*, 1995, *Journal of Neuroscience* 15:

7509-7516; and Steiner *et al.*, 1997, *Proc. National Academy of Science* 94: 2019-2024.

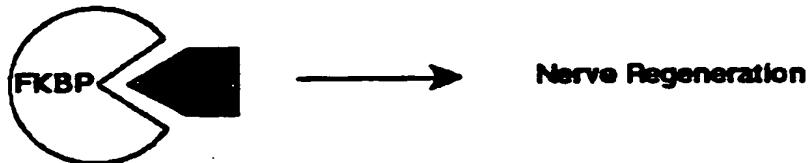
Further, the restored central and peripheral neurons appear to be functional.

Compared to protein neurotrophic molecules (BDNF, NGF, etc.), the small-molecule neurotrophins such as FK-506, FK-520, and rapamycin have different, and often advantageous, properties. First, whereas protein neurotrophins are difficult to deliver to their intended site of action and may require intra-cranial injection, the small-molecule neurotrophins display excellent bioavailability; they are active when administered subcutaneously and orally. Second, whereas protein neurotrophins show quite specific effects, the small-molecule neurotrophins show rather broad effects.

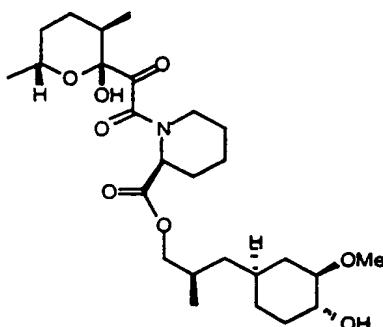
Finally, whereas protein neurotrophins often show effects on normal sensory nerves, the small-molecule neurotrophins do not induce aberrant sprouting of normal neuronal processes and seem to affect damaged nerves specifically. Neuroimmunophilin ligands have potential therapeutic utility in a variety of disorders involving nerve degeneration (e.g. multiple sclerosis, Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, peripheral neuropathies).

Recent studies have shown that the immunosuppressive and neurite outgrowth activity of FK-506, FK-520, and rapamycin can be separated; the neuroregenerative activity in the absence of immunosuppressive activity is retained by agents which bind to FKBP but not to the effector proteins calcineurin or RAPT. See Steiner *et al.*, 1997,

20. *Nature Medicine* 3: 421-428.



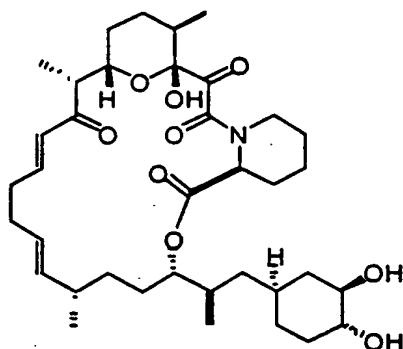
Available structure-activity data show that the important features for neurotrophic activity of rapamycin, FK-520, and FK-506 lie within the common, contiguous segments of the macrolide ring that bind to FKBP. This portion of the molecule is termed the "FKBP binding domain" (see VanDuyne *et al.*, 1993, *Journal of Molecular Biology* 229: 105-124.). Nevertheless, the effector domains of the parent macrolides contribute to conformational rigidity of the binding domain and thus indirectly contribute to FKBP binding.



"FKBP binding domain"

There are a number of other reported analogs of FK-506, FK-520, and rapamycin that bind to FKBP but not the effector protein calcineurin or RAFT. These analogs show effects on nerve regeneration without immunosuppressive effects.

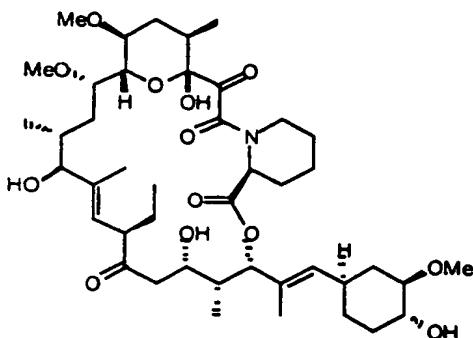
5 Naturally occurring FK-520 and FK-506 analogs include the antascomycins, which are FK-506-like macrolides that lack the functional groups of FK-506 that bind to calcineurin (see Fehr *et al.*, 1996, *The Journal of Antibiotics* 49: 230-233). These molecules bind FKBP as effectively as does FK-506; they antagonize the effects of both FK-506 and rapamycin, yet lack immunosuppressive activity.



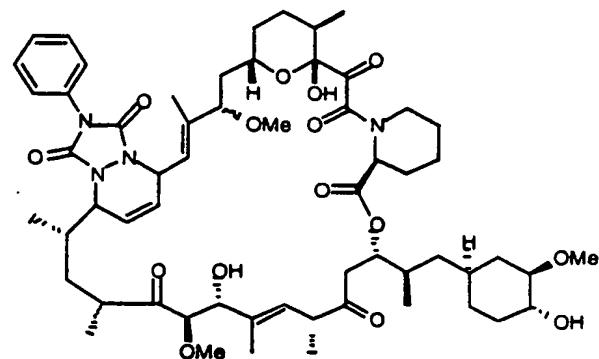
Antascomycin A

10

Other analogs can be produced by chemically modifying FK-506, FK-520, or rapamycin. One approach to obtaining neuroimmunophilin ligands is to destroy the effector binding region of FK-506, FK-520, or rapamycin by chemical modification. While the chemical modifications permitted on the parent compounds are quite limited, 15 some useful chemically modified analogs exist. The FK-520 analog L-685,818 ($ED_{50} = 0.7$ nM for FKBP binding; see Dumont *et al.*, 1992), and the rapamycin analog WAY-124,466 ($IC_{50} = 12.5$ nM; see Ocain *et al.*, 1993, *Biochemistry Biophysical Research Communications* 192: 1340-134693) are about as effective as FK-506, FK-520, and rapamycin at promoting neurite outgrowth in sensory neurons (see Steiner *et al.*, 1997).



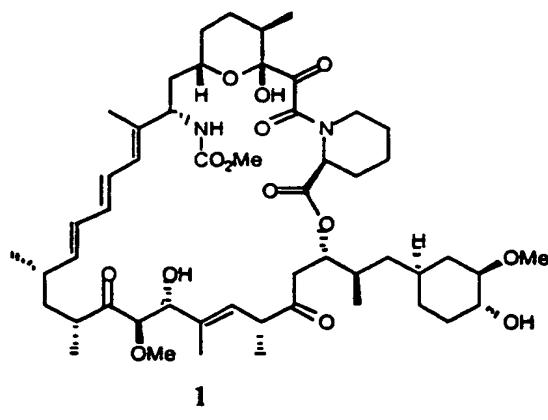
L-685,818



WAY-124,466

- One of the few positions of rapamycin that is readily amenable to chemical modification is the allylic 16-methoxy group; this reactive group is readily exchanged by acid-catalyzed nucleophilic substitution. Replacement of the 16-methoxy group of 5 rapamycin with a variety of bulky groups has produced analogs showing selective loss of immunosuppressive activity while retaining FKBP-binding (see Luengo *et al.*, 1995, *Chemistry & Biology* 2: 471-481). One of the best compounds, 1, below, shows complete loss of activity in the splenocyte proliferation assay with only a 10-fold reduction in binding to FKBP.

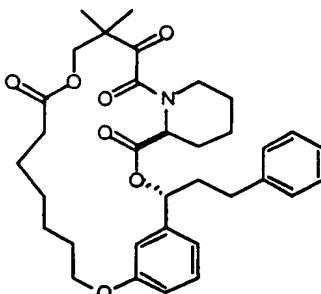
10



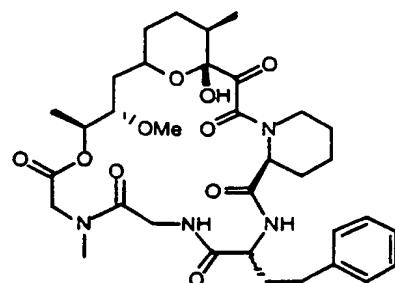
1

- There are also synthetic analogs of FKBP binding domains. These compounds reflect an approach to obtaining neuroimmunophilin ligands based on "rationally designed" molecules that retain the FKBP-binding region in an appropriate conformation 15 for binding to FKBP, but do not possess the effector binding regions. In one example, the ends of the FKBP binding domain were tethered by hydrocarbon chains (see Holt *et al.*, 1993, *Journal of the American Chemical Society* 115: 9925-9938); the best analog, 2, below, binds to FKBP about as well as FK-506. In a similar approach, the ends of the FKBP binding domain were tethered by a tripeptide to give analog 3, below, which binds

to FKBP about 20-fold poorer than FK-506. These compounds are anticipated to have neuroimmunophilin binding activity.



2



3

5 In a primate MPTP model of Parkinson's disease, administration of FKBP ligand GPI-1046 caused brain cells to regenerate and behavioral measures to improve. MPTP is a neurotoxin, which, when administered to animals, selectively damages nigral-striatal dopamine neurons in the brain, mimicking the damage caused by Parkinson's disease. Whereas, before treatment, animals were unable to use affected limbs, the FKBP ligand 10 restored the ability of animals to feed themselves and gave improvements in measures of locomotor activity, neurological outcome, and fine motor control. There were also corresponding increases in regrowth of damaged nerve terminals. These results demonstrate the utility of FKBP ligands for treatment of diseases of the CNS.

From the above description, two general approaches towards the design of non-15 immunosuppressant, neuroimmunophilin ligands can be seen. The first involves the construction of constrained cyclic analogs of FK-506 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP. The advantages of this approach are that the conformation of the analogs can be accurately modeled and predicted by computational methods, and the analogs closely resemble parent molecules that have 20 proven pharmacological properties. A disadvantage is that the difficult chemistry limits the numbers and types of compounds that can be prepared. The second approach involves the trial and error construction of acyclic analogs of the FKBP binding domain by conventional medicinal chemistry. The advantages to this approach are that the chemistry is suitable for production of the numerous compounds needed for such 25 interactive chemistry-bioassay approaches. The disadvantages are that the molecular types of compounds that have emerged have no known history of appropriate pharmacological properties, have rather labile ester functional groups, and are too conformationally mobile to allow accurate prediction of conformational properties.

The present invention provides useful methods and reagents related to the first 30 approach, but with significant advantages. The invention provides recombinant PKS

genes that produce a wide variety of polyketides that cannot otherwise be readily synthesized by chemical methodology alone. Moreover, the present invention provides polyketides that have either or both of the desired immunosuppressive and neurotrophic activities, some of which are produced only by fermentation and others of which are

5 produced by fermentation and chemical modification. Thus, in one aspect, the invention provides compounds that optimally bind to FKBP but do not bind to the effector proteins. The methods and reagents of the invention can be used to prepare numerous constrained cyclic analogs of FK-520 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP. Such compounds will show

10 neuroimmunophilin binding (neurotrophic) but not immunosuppressive effects. The invention also allows direct manipulation of FK-520 and related chemical structures *via* genetic engineering of the enzymes involved in the biosynthesis of FK-520 (as well as related compounds, such as FK-506 and rapamycin); similar chemical modifications are simply not possible because of the complexity of the structures. The invention can also

15 be used to introduce "chemical handles" into normally inert positions that permit subsequent chemical modifications.

Several general approaches to achieve the development of novel neuroimmunophilin ligands are facilitated by the methods and reagents of the present invention. One approach is to make "point mutations" of the functional groups of the

20 parent FK-520 structure that bind to the effector molecules to eliminate their binding potential. These types of structural modifications are difficult to perform by chemical modification, but can be readily accomplished with the methods and reagents of the invention.

A second, more extensive approach facilitated by the present invention is to

25 utilize molecular modeling to predict optimal structures *ab initio* that bind to FKBP but not effector molecules. Using the available X-ray crystal structure of FK-520 (or FK-506) bound to FKBP, molecular modeling can be used to predict polyketides that should optimally bind to FKBP but not calcineurin. Various macrolide structures can be generated by linking the ends of the FKBP-binding domain with "all possible"

30 polyketide chains of variable length and substitution patterns that can be prepared by genetic manipulation of the FK-520 or FK-506 PKS gene cluster in accordance with the methods of the invention. The ground state conformations of the virtual library can be determined, and compounds that possess binding domains most likely to bind well to FKBP can be prepared and tested.

Once a compound is identified in accordance with the above approaches, the invention can be used to generate a focused library of analogs around the lead candidate, to "fine tune" the compound for optimal properties. Finally, the genetic engineering methods of the invention can be directed towards producing "chemical handles" that 5 enable medicinal chemists to modify positions of the molecule previously inert to chemical modification. This opens the path to previously prohibited chemical optimization of lead compounds by time-proven approaches.

Moreover, the present invention provides polyketide compounds and the recombinant genes for the PKS enzymes that produce the compounds that have 10 significant advantages over FK-506 and FK-520 and their analogs. The metabolism and pharmacokinetics of tacrolimus has been extensively studied, and FK-520 is believed to be similar in these respects. Absorption of tacrolimus is rapid, variable, and incomplete from the gastrointestinal tract (Harrison's Principles of Internal Medicine, 14th edition, 1998, McGraw Hill, 14, 20, 21, 64-67). The mean bioavailability of the oral dosage form 15 is 27%, (range 5 to 65%). The volume of distribution (V_olD) based on plasma is 5 to 65 L per kg of body weight (L/kg), and is much higher than the V_olD based on whole blood concentrations, the difference reflecting the binding of tacrolimus to red blood cells. Whole blood concentrations may be 12 to 67 times the plasma concentrations. Protein binding is high (75 to 99%), primarily to albumin and alpha₁-acid glycoprotein. The 20 half-life for distribution is 0.9 hour; elimination is biphasic and variable: terminal-11.3 hr (range, 3.5 to 40.5 hours). The time to peak concentration is 0.5 to 4 hours after oral administration.

Tacrolimus is metabolized primarily by cytochrome P450 3A enzymes in the liver and small intestine. The drug is extensively metabolized with less than 1% excreted 25 unchanged in urine. Because hepatic dysfunction decreases clearance of tacrolimus, doses have to be reduced substantially in primary graft non-function, especially in children. In addition, drugs that induce the cytochrome P450 3A enzymes reduce tacrolimus levels, while drugs that inhibit these P450s increase tacrolimus levels.

Tacrolimus bioavailability doubles with co-administration of ketoconazole, a drug that 30 inhibits P450 3A. See, Vincent *et al.*, 1992, *In vitro* metabolism of FK-506 in rat, rabbit, and human liver microsomes: Identification of a major metabolite and of cytochrome P450 3A as the major enzymes responsible for its metabolism, *Arch. Biochem. Biophys.* 294: 454-460; Iwasaki *et al.*, 1993, Isolation, identification, and biological activities of oxidative metabolites of FK-506, a potent immunosuppressive macrolide lactone, *Drug 35 Metabolism & Disposition* 21: 971-977; Shiraga *et al.*, 1994, Metabolism of FK-506, a

- potent immunosuppressive agent, by cytochrome P450 3A enzymes in rat, dog, and human liver microsomes, *Biochem. Pharmacol.* 47: 727-735; and Iwasaki *et al.*, 1995, Further metabolism of FK-506 (Tacrolimus); Identification and biological activities of the metabolites oxidized at multiple sites of FK-506, *Drug Metabolism & Disposition* 23: 5 28-34. The cytochrome P450 3A subfamily of isozymes has been implicated as important in this degradative process.

Structures of the eight isolated metabolites formed by liver microsomes are shown in Figure 6. Four metabolites of FK-506 involve demethylation of the oxygens on carbons 13, 15, and 31, and hydroxylation of carbon 12. The 13-demethylated (hydroxy) 10 compounds undergo cyclizations of the 13-hydroxy at C-10 to give M-I, MVI and MVII, and the 12-hydroxy metabolite at C-10 to give I. Another four metabolites formed by oxidation of the four metabolites mentioned above were isolated by liver microsomes from dexamethasone treated rats. Three of these are metabolites doubly demethylated at the methoxy groups on carbons 15 and 31 (M-V), 13 and 31 (M-VI), and 13 and 15 (M- 15 VII). The fourth, M-VIII, was the metabolite produced after demethylation of the 31-methoxy group, followed by formation of a fused ring system by further oxidation. Among the eight metabolites, M-II has immunosuppressive activity comparable to that of FK-506, whereas the other metabolites exhibit weak or negligible activities. Importantly, the major metabolite of human, dog, and rat liver microsomes is the 13- 20 demethylated and cyclized FK-506 (M-I).

Thus, the major metabolism of FK-506 proceeds via 13-demethylation followed by cyclization to the inactive M-I, this representing about 90% of the metabolic products after a 10 minute incubation with liver microsomes. Analogs of tacrolimus that do not possess a C-13 methoxy group would not be susceptible to the first and most important 25 biotransformation in the destructive metabolism of tacrolimus (i.e. cyclization of 13-hydroxy to C-10). Thus, a 13-desmethoxy analog of FK-506 should have a longer half-life in the body than does FK-506. The C-13 methoxy group is believed not to be required for binding to FKBP or calcineurin. The C-13 methoxy is not present on the identical position of rapamycin, which binds to FKBP with equipotent affinity as 30 tacrolimus. Also, analysis of the 3-dimensional structure of the FKBP-tacrolimus-calcineurin complex shows that the C-13 methoxy has no interaction with FKBP and only a minor interaction with calcineurin. The present invention provides C-13- 35 desmethoxy analogs of FK-506 and FK-520, as well as the recombinant genes that encode the PKS enzymes that catalyze their synthesis and host cells that produce the compounds.

These compounds exhibit, relative to their naturally occurring counterparts, prolonged immunosuppressive action *in vivo*, thereby allowing a lower dosage and/or reduced frequency of administration. Dosing is more predictable, because the variability in FK-506 dosage is largely due to variation of metabolism rate. FK-506 levels in blood 5 can vary widely depending on interactions with drugs that induce or inhibit cytochrome P450 3A (summarized in USP Drug Information for the Health Care Professional). Of particular importance are the numerous drugs that inhibit or compete for CYP 3A, because they increase FK-506 blood levels and lead to toxicity (Prograf package insert, Fujisawa[®] US, Rev 4/97, Rec 6/97). Also important are the drugs that induce P450 3A 10 (e.g. Dexamethasone), because they decrease FK-506 blood levels and reduce efficacy. Because the major site of CYP 3A action on FK-506 is removed in the analogs provided by the present invention, those analogs are not as susceptible to drug interactions as the naturally occurring compounds.

Hyperglycemia, nephrotoxicity, and neurotoxicity are the most significant 15 adverse effects resulting from the use of FK-506 and are believed to be similar for FK-520. Because these effects appear to occur primarily by the same mechanism as the immunosuppressive action (i.e. FKBP-calcineurin interaction), the intrinsic toxicity of the desmethoxy analogs may be similar to FK-506. However, toxicity of FK-506 is dose related and correlates with high blood levels of the drug (Prograf package insert, 20 Fujisawa[®] US, Rev 4/97, Rec 6/97). Because the levels of the compounds provided by the present invention should be more controllable, the incidence of toxicity should be significantly decreased with the 13-desmethoxy analogs. Some reports show that certain FK-506 metabolites are more toxic than FK-506 itself, and this provides an additional 25 reason to expect that a CYP 3A resistant analog can have lower toxicity and a higher therapeutic index.

Thus, the present invention provides novel compounds related in structure to FK-506 and FK-520 but with improved properties. The invention also provides methods for making these compounds by fermentation of recombinant host cells, as well as the recombinant host cells, the recombinant vectors in those host cells, and the recombinant 30 proteins encoded by those vectors. The present invention also provides other valuable materials useful in the construction of these recombinant vectors that have many other important applications as well. In particular, the present invention provides the FK-520 PKS genes, as well as certain genes involved in the biosynthesis of FK-520 in recombinant form.

FK-520 is produced at relatively low levels in the naturally occurring cells, *Streptomyces hygroscopicus* var. *ascomyceticus*, in which it was first identified. Thus, another benefit provided by the recombinant FK-520 PKS and related genes of the present invention is the ability to produce FK-520 in greater quantities in the 5 recombinant host cells provided by the invention. The invention also provides methods for making novel FK-520 analogs, in addition to the desmethoxy analogs described above, and derivatives in recombinant host cells of any origin.

The biosynthesis of FK-520 involves the action of several enzymes. The FK-520 PKS enzyme, which is composed of the *fkbA*, *fkbB*, *fkbC*, and *fkbP* gene products, 10 synthesizes the core structure of the molecule. There is also a hydroxylation at C-9 mediated by the P450 hydroxylase that is the *fkbD* gene product and that is oxidized by the *fkbO* gene product to result in the formation of a keto group at C-9. There is also a methylation at C-31 that is mediated by an O-methyltransferase that is the *fkbM* gene product. There are also methylations at the C-13 and C-15 positions by a 15 methyltransferase believed to be encoded by the *fkbG* gene; this methyltransferase may act on the hydroxymalonyl CoA substrates prior to binding of the substrate to the AT domains of the PKS during polyketide synthesis. The present invention provides the genes encoding these enzymes in recombinant form. The invention also provides the genes encoding the enzymes involved in ethylmalonyl CoA and 2-hydroxymalonyl CoA 20 biosynthesis in recombinant form. Moreover, the invention provides *Streptomyces hygroscopicus* var. *ascomyceticus* recombinant host cells lacking one or more of these genes that are useful in the production of useful compounds.

The cells are useful in production in a variety of ways. First, certain cells make a useful FK-520-related compound merely as a result of inactivation of one or more of the 25 FK-520 biosynthesis genes. Thus, by inactivating the C-31 O-methyltransferase gene in *Streptomyces hygroscopicus* var. *ascomyceticus*, one creates a host cell that makes a desmethyl (at C-31) derivative of FK-520. Second, other cells of the invention are unable to make FK-520 or FK-520 related compounds due to an inactivation of one or more of the PKS genes. These cells are useful in the production of other polyketides produced by 30 PKS enzymes that are encoded on recombinant expression vectors and introduced into the host cell.

Moreover, if only one PKS gene is inactivated, the ability to produce FK-520 or an FK-520 derivative compound is restored by introduction of a recombinant expression vector that contains the functional gene in a modified or unmodified form. The 35 introduced gene produces a gene product that, together with the other endogenous and

functional gene products, produces the desired compound. This methodology enables one to produce FK-520 derivative compounds without requiring that all of the genes for the PKS enzyme be present on one or more expression vectors. Additional applications and benefits of such cells and methodology will be readily apparent to those of skill in the art after consideration of how the recombinant genes were isolated and employed in the construction of the compounds of the invention.

5 The FK-520 biosynthetic genes were isolated by the following procedure. Genomic DNA was isolated from *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) using the lysozyme/proteinase K protocol described in Genetic 10 Manipulation of *Streptomyces* - A Laboratory Manual (Hopwood *et al.*, 1986). The average size of the DNA was estimated to be between 80 - 120 kb by electrophoresis on 0.3% agarose gels. A library was constructed in the SuperCos™ vector according to the manufacturer's instructions and with the reagents provided in the commercially available kit (Stratagene). Briefly, 100 µg of genomic DNA was partially digested with 4 units of 15 *Sau3A* I for 20 min. in a reaction volume of 1 mL, and the fragments were dephosphorylated and ligated to SuperCos vector arms. The ligated DNA was packaged and used to infect log-stage XL1-BlueMR cells. A library of about 10,000 independent cosmid clones was obtained.

Based on recently published sequence from the FK-506 cluster (Motamedi and 20 Shafiee, 1998, *Eur. J. Biochem.* 256: 528), a probe for the *fkbO* gene was isolated from ATCC 14891 using PCR with degenerate primers. With this probe, a cosmid designated pKOS034-124 was isolated from the library. With probes made from the ends of cosmid pKOS034-124, an additional cosmid designated pKOS034-120 was isolated. These 25 cosmids (pKOS034-124 and pKOS034-120) were shown to contain DNA inserts that overlap with one another. Initial sequence data from these two cosmids generated sequences similar to sequences from the FK-506 and rapamycin clusters, indicating that the inserts were from the FK-520 PKS gene cluster. Two *Eco*RI fragments were subcloned from cosmids pKOS034-124 and pKOS034-120. These subclones were used 30 to prepare shotgun libraries by partial digestion with *Sau*3AI, gel purification of fragments between 1.5 kb and 3 kb in size, and ligation into the pLitmus28 vector (New England Biolabs). These libraries were sequenced using dye terminators on a Beckmann CEQ2000 capillary electrophoresis sequencer, according to the manufacturer's protocols.

35 To obtain cosmids containing sequence on the left and right sides of the sequenced region described above, a new cosmid library of ATCC 14891 DNA was prepared essentially as described above. This new library was screened with a new *fkbM*

probe isolated using DNA from ATCC 14891. A probe representing the *fkbP* gene at the end of cosmid pKOS034-124 was also used. Several additional cosmids to the right of the previously sequenced region were identified. Cosmids pKOS065-C31 and pKOS065-C3 were identified and then mapped with restriction enzymes. Initial sequences from 5 these cosmids were consistent with the expected organization of the cluster in this region. More extensive sequencing showed that both cosmids contained in addition to the desired sequences, other sequences not contiguous to the desired sequences on the host cell chromosomal DNA. Probing of additional cosmid libraries identified two additional cosmids, pKOS065-M27 and pKOS065-M21, that contained the desired sequences in a 10 contiguous segment of chromosomal DNA. Cosmids pKOS034-124, pKOS034-120, pKOS065-M27, and pKOS065-M21 have been deposited with the American Type Culture Collection, Manassas, VA, USA. The complete nucleotide sequence of the coding sequences of the genes that encode the proteins of the FK-520 PKS are shown 15 below but can also be determined from the cosmids of the invention deposited with the ATCC using standard methodology.

Referring to Figures 1 and 3, the FK-520 PKS gene cluster is composed of four open reading frames designated *fkbB*, *fkbC*, *fkbA*, and *fkbP*. The *fkbB* open reading frame encodes the loading module and the first four extender modules of the PKS. The *fkbC* open reading frame encodes extender modules five and six of the PKS. The *fkbA* open 20 reading frame encodes extender modules seven, eight, nine, and ten of the PKS. The *fkbP* open reading frame encodes the NRPS of the PKS. Each of these genes can be isolated from the cosmids of the invention described above. The DNA sequences of these genes are provided below preceded by the following table identifying the start and stop codons of the open reading frames of each gene and the modules and domains contained 25 therein.

	<u>Nucleotides</u>	<u>Gene or Domain</u>
	complement (412 - 1836)	<i>fkbW</i>
	complement (2020 - 3579)	<i>fkbV</i>
30	complement (3969 - 4496)	<i>fkbR2</i>
	complement (4595 - 5488)	<i>fkbR1</i>
	5601 - 6818	<i>fkbE</i>
	6808 - 8052	<i>fkbF</i>
	8156 - 8824	<i>fkbG</i>
35	complement (9122 - 9883)	<i>fkbH</i>
	complement (9894 - 10994)	<i>fkbI</i>
	complement (10987 - 11247)	<i>fkbJ</i>
	complement (11244 - 12092)	<i>fkbK</i>
	complement (12113 - 13150)	<i>fkbL</i>
40	complement (13212 - 23988)	<i>fkbC</i>

	complement (23992 - 46573)	<i>fkbB</i>
	46754 - 47788	<i>fkbO</i>
	47785 - 52272	<i>fkbP</i>
	52275 - 71465	<i>fkbA</i>
5	71462 - 72628	<i>fkbD</i>
	72625 - 73407	<i>fkbM</i>
	complement (73460 - 76202)	<i>fkbN</i>
	complement (76336 - 77080)	<i>fkbQ</i>
	complement (77076 - 77535)	<i>fkbS</i>
10	complement (44974 - 46573)	CoA ligase of loading domain
	complement (43777 - 44629)	ER of loading domain
	complement (43144 - 43660)	ACP of loading domain
	complement (41842 - 43093)	KS of extender module 1 (KS1)
	complement (40609 - 41842)	AT1
15	complement (39442 - 40609)	DH1
	complement (38677 - 39307)	KR1
	complement (38371 - 38581)	ACP1
	complement (37145 - 38296)	KS2
	complement (35749 - 37144)	AT2
20	complement (34606 - 35749)	DH2 (inactive)
	complement (33823 - 34480)	KR2
	complement (33505 - 33715)	ACP2
	complement (32185 - 33439)	KS3
	complement (31018 - 32185)	AT3
25	complement (29869 - 31018)	DH3 (inactive)
	complement (29092 - 29740)	KR3
	complement (28750 - 28960)	ACP3
	complement (27430 - 28684)	KS4
	complement (26146 - 27430)	AT4
30	complement (24997 - 26146)	DH4 (inactive)
	complement (24163 - 24373)	ACP4
	complement (22653 - 23892)	KS5
	complement (21420 - 22653)	AT5
	complement (20241 - 21420)	DHS
35	complement (19464 - 20097)	KR5
	complement (19116 - 19326)	ACPS
	complement (17820 - 19053)	KS6
	complement (16587 - 17820)	AT6
	complement (15438 - 16587)	DH6
40	complement (14517 - 15294)	ER6
	complement (13761 - 14394)	KR6
	complement (13452 - 13662)	ACP6
	5236? - 53576	KS7
	53577 - 54716	AT7
45	54717 - 55871	DH7
	56019 - 56819	ER7
	56943 - 57575	KR7
	57710 - 57920	ACP7
	57990 - 59243	KS8
50	59244 - 60398	AT8
	60399 - 61412	DH8 (inactive)
	61548 - 62180	KR8

	62328 - 62537	ACP8
	62598 - 63854	KS9
	63855 - 65084	AT9
	65085 - 66254	DH9
5	66399 - 67175	ER9
	67299 - 67931	KR9
	68094 - 68303	ACP9
	68397 - 69653	KS10
	69654 - 70985	AT10
10	71064 - 71273	ACP10

1' GATCTCAGGC ATGAAGTCCT CCAGGCGAGG CGCCGAGGTG GTGAACACCT CGCCGCTGCT
 61 TGTACGGACC ACTTCAGTCA GCGGCGATTG CGGAACCAAG TCATCCGAA TAAAGGGCGG
 121 TTACAAGATC CTCACATTGC GCGACCGCCA GCATACGCTG AGTTGCCTCA GAGGCAAACC
 15 181 GAAAGGGCGC GGGCGGTCCG CACCAGGGCG GAGTACGCGA CGAGAGTGGC GCACCCGCGC
 241 ACCGTACACT CTCTCCCCCG CGCGCGGGAT GCCCAGGTG ACACGGTTGG GCTCTCCTCG
 301 ACGCTGAACA CCCGCGCGT GTGGCGTCGG GGACACCGCC TGGCATCGGC CGGGTGACGG
 361 TACGGGGAGG GCGTACGGCG GCCGTGGCTC GTGCTCACGG CCGCCGGGGCG GTCATCCGTC
 421 GAGACGGCAC TCGGCAGCA GGGACGCCTG GTCGGCACCT GCGGGCCGGA CGACCGTGTG
 20 481 GTTCGCGGGC GGGCGGTGGC CGGTGGTGGAG CCAGCTCTCC AGGGCGGTGA AGGCTGAGCG
 541 GTGACACGGC AGCAAAGGC GGAGTCGGTC GGGGAAGGTG TCGACGAGGG CGTGGTGTG
 601 CGTGCCTGTC TCGATGCGGT AGTAGCGGTA CGGCGCGCCA GGCCGCTGCC GGACATACGC
 661 GCGTACACGT CGGAGCCCGG CGGGCAGGCA GCAGCACGTC GAGAGTGTCT GGATGGTGTAT
 721 CAGCGGTTG CCGATACGAC CGGTCAACGC GATGCGTTCC ACGGCCGCGT GGACCGCUGGA
 25 781 GGAGCGGGTG CGCTAGTCGT AGTCGGCATC GCAGCCCGGG ACCGTCCTCCG GGGCGCAATA
 841 CGGTGFGCCG GCTTCCTCTT CCCCATCGAA GCCGGGGTCG AACTCCTCGC GGTAGACGCG
 901 CTGGTCAAGA TCCCAGTAGA CCTCGTGGTG GTACGCCAC AAGAACCTGG AGTCGGCCGG
 961 GAACCCGGCG CGGAGCAGCG CCTCGCGCGC CTGGCCGGCT GCGGGGGCCG CTGCGCGTA
 30 1021 GGTGGGGTAG TCGCGCAGGG CGGCCGGCAG GAAGGTGAAG AGGTTGGGAC CCTCCGCGCG
 1081 CCACAGGGTG CCTTCCCAGT CGACTCCTCC GTCGTACAGC TCGGGATGGT TCTCCAGCTG
 1141 CCAGCGCACG AGGTAGCGC CGTTGGACAT CCCGGTGACC AGGGTGCCT CGAGCCGGCG
 1201 GTGGTAGCGC TGGGCACCG ACGGCGGGGGC GGGCCGGTC AGCTGGTGA GCGGGTGTGTT
 1261 CCACCTCGCGC ACGGCGTCGC CGGGCCGGGA GCCATCACGG TAGAACCGGG GGCGGGTGTGTT
 1321 GCCCTTGTGCG GTGGCGCGT AGGCATAACC GCGGGCGAGC ACCCAGTCGG CGATGGCCCG
 35 1381 GTCGTTGGCG TACTGCTCGC GGTACCGGG GGTGCGGCC ACGACCAGGC CACCGTTCCA
 1441 GCGGTCGGGC AGCGGATGA CGAACCTGGGC GTCGTTGGTCA CACCCGTGGT TGGTGTGTT
 1501 GGTGGAGGTG TCGGGGAAGT AGCCGTCGAT CTGGATCCCG GGCACCTCCGG TGGGAGTGGC
 1561 CAGGTTCTTG GGCGTCAGCC CTGCCCAGTC CGCCGGGTG GTGTGGCCGG TGGCCGCCGT
 1621 TCCCGCCGTG GTCAGCTCGT CCAGGCACTC GGCCTGCTGA CGTGCCTGCC CGGGACACG
 40 1681 CAGCTGGGAC AGACGGGCGC AGTGACCGTC CGGGGCATCG GGAGCAGGGC GGGCCGTGGC
 1741 CGGTGAGGGG AGCAGGACGG CGACTGCGGC CAGGGTGAGA GCGCCGAGGC CGGTGCGTCT
 1801 TCTCGGGGCC CGTCCGACAC CGAGGGGCAG AACCATGGAG AGCCTCCAGA CGTGCAGATG
 1861 GATGACGGAC TGGAGGCTAG GTCGCGCAGC GTGGAGACGA ACATGGGTGC GCCCCTCATG
 1921 ACTGAGGCCC CTCAGAGGTG GGGCGCCGCC ATGACGGGCG CGGGACCGCG GGCCTCCGG
 45 1981 GGCGGFGCCC CGGGCGCCA CGGGTCCGG GTCGGGGGT CAGGGACAGG TGTGTTCGC
 2041 GACGGTGAAG TAGCCGGTCG GCGACTCTTT CAAGGTGGTC GTGACGAAGG TGGTGTACAG
 2101 GCCCATGTTC TGGCGGAGC CCTTGGCGTA GGTGTAACCG GCGCTGTCG TGGCGCGGCC
 2161 CGCTTGGACG TGAGCGTAGT TGCCGGCGGT CCAGCAGACG GCGTGGCAC CGGTGCTCTG
 2221 CGCGGTGACC GCGCCCGAGA GCGGTCCGGC CTTGCGTCC CGTCCCAGGG CGGCACCGC
 50 2281 GTAGGTGTGC GATGTGCCCG CCCTCAGGCC GGTGTCCTG TACGACGTCG TGGCGGACGT
 2341 GGTGATCTGG GCACCGTCGC GGTGGACGGC GTAGTCGGT GCGCCGTCG CGGGTTCCA
 2401 GGTCAAGGCTG ATGGTGGTGT CGGTGGCGCC GGTGGCGGC AGGCCGGACG GAGCGGGCAG
 2461 CGAACCGGGG TCGGAGGCGG ATCCGCTCAG GCGGAAGAAC TCGTGTATCC AGTAGCTGGA
 2521 ACAGATCGAG TCCAGGAAGT AGGCAGGCGCC GGTGCTGCC CACTGCTGTG CTCCGGTGC
 55 2581 GGGATCGACC GGGGTGCCGT GCCCGATGCC CGGCACCCGG TTCACCTCCA CGGCCACCGA
 2641 TCCGTCCGCC GCCAGGTACT CCTCGTGCCG GGTGGAGTTC GGGCGATCA CCGAGGTACG
 2701 GTCCGGCGTC TGGGACACGC CGTGCACAGC GGTCCACTGG TCGCGCAACT CGTCCGGCGT
 2761 GCGCGGCCGCG ACGGTGGTGT CCTTGTGCCG GTGCCAGATG GCCACGCCGG GCCACGGGCC
 2821 CGACCAACGAG GGGTAGCCGT CACGGACCCG CGCGCCCCAC TGGTCCGCCGG TCAGGTGGT
 60 2881 CCCGGGGTTC ATGCACAGGT ACAGCGCTGCT GACGTCGGT GCACAGCCG AAGGGCAGGCC
 2941 GCGGACGACC GCGCCGGCCT GGAAGACGTC CGGATAGGTG CGGAGCATCA CCGACGTCA

3001	GGCACCGCCG	GCGGACAGCC	CGGTGATGTA	GGTGCCTGG	GGTCGCCGC	CGTAGGCCGA	
3061	GACGGTGTGA	GCGGCCATCT	GCCGGATCGA	CGCGCTTCG	CCCTGGCCC	TGCGTTGTC	
3121	GCTGCTCTGG	AACCAGTTGA	AGCACCTGT	CGCGTGTTC	GACGACGTGG	TCTCGCGAA	
3181	CACGAGCAGG	AAGCCATAGC	GGTCGCCGAA	TGAGAGCAGG	CCGGAGTTG	CGGCGTAGCC	
5	3241	CTGGCGTCC	TGGGTGCAAC	CGTGCAGGGC	GAACACCACC	GCCGGCTCCG	CGGGCAGGGA
3301	CGGGGCCGG	TAGACGTACA	TGTCAGCCG	GCCCAGGTT	GTGCCGAAGT	CCGCGACCTC	
3361	GGTCAGGTCC	GCCTTGGTCA	GACCGGGCT	GGCCAGGCC	GCCGCCGCGT	GGGCCGTCGG	
3421	CGCCGGGCCG	AGCAGGGCCG	CTCCGAGTAC	GAGGGCCACG	ACGGCCACGA	GACGGGAGG	
10	3481	CACCCCCCGC	CGTCCCGGAC	GCGACAACGA	CCCGACCGGC	GGCGAGGAGG	AGAGGGGAA
3541	CAGCGGGGTG	AGGATTCCCC	GGAACGGCGG	CGGCTGCATG	GGGGCTCCCT	CGATGTCGTG	
3601	GGGGGGACAC	GGAGGGCTCC	CTGACGTCGA	TCAGTGGGAG	CGCCCCGGTG	CCCGCACCAG	
3661	TAGGGGTGGT	TCAACCCGCA	ACGGTATGGC	CCGGAGCACC	ACACCCCGCA	CCGCGCAGATG	
3721	TGCGCCCGGA	CGGATTGTGT	CGCCTTGCGG	AATCTGATAC	CCGGACGCGA	CGAACGCC	
15	3781	ACCCGACACG	GGTAGGGCGT	CATGGTGTCC	GACTCGGCCG	GTCGGCTTG	CTGCCCTGG
3841	ACGGACCGGG	CGTCGGCGA	CCGGCGCTCG	GGGGCTGGG	GGTATGGCG	GCGGAGGACG	
3901	CCAGCCGCGT	GGGGCGGCCG	CGCCCAAGTG	CAGTACGCCG	ACCGTGGCCG	GGGGAGGGC	
3961	CGGA.CCGGTC	AGTGCAGTCC	CGCGGCCCTG	CGGGACCGCT	CGTCCCAGAC	GGGTTCCACC	
4021	GCGGCGAAC	GGGGTCCCGT	TCCCGCGCGG	TAGACCATCA	GTGTCCGCTC	GAAGGTGATG	
20	4081	ACGATGACAC	CGTCCTGGTT	GTAGCCGATG	GTGCGCACGC	TGATGATGCC	TACGTCAGGT
4141	CGGCTGGCGG	ACTCCCAGG	GTTCAGGACC	TCGGACTGCG	AGTAGATGGT	GTCGCCCTCG	
4201	AAGACCGGGT	TCGGCAGCCT	GACCCGGTCC	CAGCCGAGGT	TGGCCATCAC	ATGCTGGGAG	
4261	ATGTCGGTGA	CGCTCTGCC	GGTGACCAGG	GCGAGGGTGA	AGGTGGAGTC	CACCAAGCGC	
4321	TTGCCCGCAGG	TGGTCCCGC	CGAGTAGTGTG	CGGTCGAAGT	GCAGCGGCCG	GGTGTCTG	
25	4381	GTCAGGAGCG	TGAGCCAGGA	GTGTCGGT	TCCAGGACCG	TGCGGCCAG	GGGGTGGCGG
4441	TACACGTCG	CGGTGGTGA	GTCTCGAAG	TAGCGCCCT	GCCAGCCCTC	GACCAACAGCG	
4501	GTGCGGGTGG	CGTCCTGGTC	CGGGTTCTCA	GTCGTATGG	CGCTCATTCT	GGGAAGTCCC	
4561	CGGTCGGCTG	TGAAATGCCG	AACTTCACC	GGGCTCATAC	GTGCGGCCGA	TGAGCCCTGG	
4621	ACCGTACGTA	GTCGTAGAAC	CTCGCCACCA	CTGGCGCGC	TGGTCTCCG	GCGAGTGTGA	
30	4681	CCACGCGAC	CGTGCGCCG	GCCTGCGGGT	CGTCGAGCG	CACGGCGACG	GCGTGGTCAC
4741	CGGGCCCGGA	CGGGCTGCCG	GTGAGGGGGG	CGACGCCAC	ACCGAGGCCG	GGGGCGACCA	
4801	GGGGCCCGCAG	CGTGCTCAGC	TCGGTGCTCT	CCAGGACGAC	CCGGGGCACG	AATCCGGCC	
4861	CGGGCGACAG	CGGGTCCGGT	ATCTGGCGCA	GTCCGAAGAC	CGGCTCCAGT	GCCACGAACG	
4921	CCTCATCGGC	CAGCTCCCG	GTCCGCACCC	GGCGCGTCT	GGCCAGCCG	TGTCCGGGTG	
4981	GGACGAGCAG	GCACAGTCC	TCGTCCCGCA	GTGGTGTCCA	CTCCACATCG	TCCCCGGCGG	
35	5041	GTGCGGGCT	GGTCAGCCCC	AGGTCCAGCC	TGCTGTTGCG	GACGTCGTG	ACCACGGCGT
5101	CGGCGCGTC	GCGCGCAGT	TCGAAGGTGG	TGCCGGGAGC	CAGCCGGCGG	TACCCGGCGA	
5161	GGAGGTCGGG	CACCAGCCAG	GTGCCGTAGG	AGTGCAGGAA	ACCCAGTGCC	ACGGTGCCGG	
5221	TGTCGGGGTC	GATCAGGGCG	GTGATGCGCT	GTCGGCGCC	GGAGACCTCA	CTGATCGCGC	
40	5281	GCAGGGCGTG	GGCGCGGAAG	ACCTCGCCGT	ACTTGGTGA	CCGGAGCCGG	TTCTGGTGC
5341	GGTCGAACAG	CGGCACGCC	ACTCGTCGCT	CCAGCCGCC	GATGGCCCTG	GACAGGGTCG	
5401	GCTGGGAGAT	GTTGAGCCGT	TCCCGGGTGA	TCGTACGTG	CTCGTGTCTG	GCAAGGCCG	
5461	TGAACCACTG	CAACTCCCCT	ATCTCCATGC	AGGGACTATA	CGTACCGGGC	ATGGCTCTGG	
5521	CGAGGTTTCG	TCATTTACA	GCGGCCGGG	GGCGGCCAC	AGTGAGTCCT	CACCAACCAG	
45	5581	GACCCATGG	GAGGGACCCC	ATGTCCGAGC	CGCATCCTCG	CCCTGAACAG	GAACGCCCG
5641	CGGGCCCGCT	GTCCGGCTG	CTCGTGGTT	CTTGAGGCA	GGCCGTGCG	GCTCCGTTG	
5701	CCACCCGCCA	CCTGGCGAC	CTGGCGCC	GTGTCATCAA	GATGAAACG	CCCGCGAGCG	
5761	GCGACCTCGC	CCGCGGCTAC	GACCGCACGG	TGCGTGGCAT	GTCCAGCCAC	TTCGTCCTG	
5821	TGAACGGGG	GAAGGAGAGC	GTCCAGCTCG	ATGTGCGCTC	GCCGGAGGGC	AACCGGACCC	
5881	TGCAACGCC	GGTGGACCGG	GCCGATGTC	TGGTGCAGAA	TCTGGCACCC	GGCGCCGCGG	
50	5941	GCCGCTGGC	ATCGGCACCC	AGGTCTCGC	GCGGAGCCAC	CGAGGCTGAT	CACCTGCGG
6001	CATATCCGGC	TACGGCAGTA	CCGGCTGCTA	CCGGCACCC	CAAGGCGTAC	GACCTCTGG	
6061	TCCAGTGC	AGCGGGCGT	GTCTCCATCA	CCGGCACCC	CGAGGACCCC	TCCAAGGTGG	
6121	GCCTGTCCAT	CGCGGACATC	TGTCGGGGGA	TGTACGCGT	CTCCGGCATC	CTCACGGCC	
6181	TGCTGAAGCG	GGCCCGCACC	GGCCGGGGCT	CGCAGTTGGA	GTTCTCGATG	CTCGAAGGCC	
55	6241	TGGTGAATG	GATGGGATAC	GCGAGTACT	ACACGGCTA	GGCGGGCAC	GCTCCGGCCC
6301	GCGCCGGCG	CAGCCACCG	ACGATGCC	CCTACGGCCC	TTTCACCC	CGCGACGGGC	
6361	AGACGATCAA	TCTCGGGCTC	CAGAACGAGC	GGGAGGGG	TTCTTCTG	GGTGTGCG	
6421	TACAACGCC	CGGTCTCTG	GACGACCCGC	GCTTTCCGG	CAACGCCGAC	GGGGTGGCGC	
6481	ACCGCACCGA	GCTCGACGCC	CTGGTGAGCG	AGGTGACGGG	CACGCTCACC	GGCGAGGAAC	
60	6541	TGGTGGCGC	GCTGGAGGAG	GGCTCGATCG	CCTACGCACG	CCAGCGCAC	GTGCGGGAGT
6601	TCAGCGAAC	CCCCCAACTG	CGTGACCGTG	GACGCTGGC	TCCGTTGCGAC	AGCCCGGTCG	
6661	GTGCGCTGGA	GGGCCTGATC	CCCCCGGTCA	CCTTCACGG	CGAGCACCCG	GGCGGGCTGG	
6721	GCCGGGTCCC	GGAGCTGGC	GAGCATACCG	AGTCCGTCCT	GGCGTGGCTG	GCCGCGCCCC	
6781	ACAGCGCCGA	CCGCGAAGAG	GCCGGCCATG	CCGAATGAAC	TCACCCGGAGT	CCTGATCCTG	

6841	GCCGCCGTGT	TCCTGCTCGC	CGGCGTACGG	GGGCTGAACA	TGGGCCTGCT	CGCGCTGGTC
6901	GCCACCTTTC	TGCTCGGGGT	GGTCGCACTC	GACCGAACGC	CGGACGAGGT	GCTGGCGGGT
6961	TTCCCCGCGA	GCATTTCT	GGTGCCTGGTC	GCCGTCACGT	TCCTCTTCGG	GATCGCCCGC
7021	GTCAACGGCA	CGGTGGACTG	GCTGGTACGT	GTCGCCTGTC	GGGCGGTGGG	GGCCCGGGTG
7081	GGAGCCGTCC	CCTGGGTGCT	CTTCGGCCTG	GCGGCACTGC	TCTGCGCGAC	AGGCGCGGCC
7141	TCGCCCGCGG	CGGTGGCGAT	CGTGGCGCCG	ATCAGCGTCG	CGTTCGCCGT	CAGGCACCGC
7201	ATCGATCCGC	TGTAACGCCG	ACTGATGGCG	GTGAAACGGGG	CCGCAGCCGG	CAGTTTCGCC
7261	CCCTCCGGGA	TCCTGGCGG	CATCGTCCAC	TCGGCGCTGG	AGAAGAACCA	TCTGCCCGTC
7321	AGCGGGGGGC	TGCTCTTCGC	AGGCACCTTC	GCCTTCAACC	TGGCGGTGCG	CGCGTJTC
7381	TGGCTCGTCC	TCGGGCCGAG	GCGCCTCGAA	CCACATGACC	TGGACGAGGA	CACCGATCCC
7441	ACCGAAGGGG	ACCCGGCTTC	CCGCCCCCGGC	GCGGAACACG	TGATGACGCT	GACCGCGATG
7501	GCCCGCGTGG	TGCTGGGAAC	CACGGTCTCT	TCCCTGGACA	CCGGCTTCTC	GGCCCTCAC
7561	TTGGCGCGT	TGCTGGCGCT	GCTCTTCCC	CGCACCTCCC	AGCAGGCCAC	CAAGGAGATC
7621	GCCTGGGGCC	TGGTGGTGC	GGTATGGGGG	ATCGTGCACCT	ACGTCGCCCT	GCTCAGGAG
7681	CTGGGCATCG	TGGACTCCCT	GGGAAAGATG	ATCGCGCGA	TCGGCACCCC	GCTGCTGGCC
7741	GCCCTCGTGA	TCTGCTACGT	GGCGGGTGT	GTCTCGGCCT	TGCGCTCGAC	CACCGGGATC
7801	CTCGGTGCC	TGATGCCGCT	GTCCGAGCCG	TTCCTGAAGT	CCGGTGCCT	CGGGACGACC
7861	GGCATGGTGA	TGGCCCTGGC	GGCCGCGGGC	ACCGTGGTGG	ACCGGAGTCC	CTTCTCCACC
7921	AATGGTGTC	TGGTGGTGGC	CAACGCTCCC	GAGCGCTGC	GGGCCCGGGT	GTACCAAGGGG
7981	TTGCTGTGGT	GGGGCGCCGG	GGTGTGCGCA	CTGGCTCCC	GGGCCGCCCTG	GGCGGCCCTC
8041	GTGGTGGCGT	GAGCGCAGCG	GAGCGGGAA	CCCCTGGAGC	CCGTTTCCC	TGCTGTGCG
8101	CTGACGTAGC	GTCAAGTCCA	CGTGCCGGGC	GGGCAGTACG	CCTAGCATGT	CGGGCATGGC
8161	TAATCAGATA	ACCCGTCCG	ACACGCTGCT	CGCTTACGTA	CGGAAGGGT	CCCTGCGC
8221	TGACGAGGTC	CTGAGCCGGC	TGCGCGCGA	GACGGCCGAG	CTGCCGGGGC	GTGGCGTACT
8281	GCCGGTGCAG	GCCGAGGAGG	GACAGTCCCT	CGAGTTCCTG	GTGCGGTTGA	CCGGCGCGC
8341	TCAGGTGCTG	GAGATCGGG	CGTACACCGG	CTACAGCACG	CTCTGCC	CCCAGGATT
8401	GGCGCCCGGG	GGCCGTGTGG	TGACGTGCGA	TCTCATGCC	AAAGTGGCCG	AGGTGGCGA
8461	GCGGTACTGG	GAGGAGGCCG	GGGTTGCCGA	CCGGATCGAC	GTCCGGATCG	GCGACGUCCG
8521	GACCGTCCTC	ACCGGGCTGC	TCGACGAGGC	GGGCGGGGG	CCGGAGTCGT	TCGACATGGT
8581	GTTCATCGAC	GCCGACAAGG	CCGGCTACCC	CGCCTACTAC	GAGGCGGGC	TGCCGCTGGT
8641	ACGCCGCGGC	GGGCTGATCG	TCGTCGACAA	CACGCTGTT	TTCGGCCGGG	TGCCGACGA
8701	AGCGGTGCAG	GACCCGGACA	CGTCGCGGT	ACGCGAACTC	AACGCGCAC	TGCGCGACGA
8761	CGACCGGGTG	GACCTGGCA	TGCTGACGAC	GGCCGACGGC	GTCACCTGC	TGCGAAACG
8821	GTGACCGGGG	CGATGTCGGC	GGCGGTCA	GTCAGCGTC	TGGCGCGGG	CTCGCGGAG
8881	GGCTCCAGAT	GCAGGCGTTC	GACGCCGGC	GCGGAAGCGC	CCGCCACCTC	GGACACGCG
8941	GGGCACTCGG	AGTCCCGAA	GCCCAGCAAC	CGGTAGGCGA	TCTCCATCAT	CGGGTTGCC
9001	TCCGTACGCC	GGAAAGTCCGC	CACCAGGTGC	GCCCCCGC	GGGCGCCCTG	GTCCGTGAGC
9061	CAGTTCAAGGA	TCGTCGACCC	GGCACCGAAC	GACACGACCC	GGCAGGACGT	GGCGAGCAGT
9121	TTCAGGTGCC	ACGTCGACGG	CTTCTTCTC	AGCAGGATGA	TGCGGACGGC	GGCGTGC
9181	CCGAAGCGGT	CGCCCCATGGT	GACGACGAGG	ACCTCATGGG	CGGGATCGGT	GAGCACGCGC
9241	GCAGGTGCGG	GTCGGAGTAG	TGACGCGCCG	TCGCGTTCAT	CTGGCTGGTC	CGCAGCGTCA
9301	GTTCCTCGAC	GCGGTGAGT	TCCCTCTCCC	CCGCGGTG	GATCGTCATG	GAGAGGTCGA
9361	GCGAGCGCAG	GAAGTCTCG	TGGGACCGG	AGTACGCC	CCGGGCTGG	TGCGCGCGA
9421	AACCCGCTG	GTACATCAGG	CGGCGCCGAC	GCGAGTCGAC	CGTGGACACC	GGCGGGCTGA
9481	ACTCCGGCAG	CGACAGGAGC	GTGGCGCCT	GTCGCC	GTAGCACCGC	ACCTCGGGCA
9541	GGTGGAACGC	CACCTGGCA	CGCTCGGCG	GCTGGTCGTC	GATGAACCGC	ATCGTGGTCG
9601	GTGCGAACGT	CAGCTCCG	GGCATCTCG	GGACGGACTG	CGACTTCG	CCCCATCCGA
9661	TGCGGGCCAG	CACGAAGTAC	TCCGCCACAC	CGAGGCGTTC	CAGACGCTCC	CACCGAGGT
9721	CGTGGTCGTT	CTTGCTCGCC	ACCGCCTGGA	GGATGCCGCG	GTCGTGAGC	GTGGTATCA
9781	CCTCGCGGAT	CTCGTCCG	AGGACCA	CGTCGTC	CAGCACGGT	CCCCGCCACA
9841	AGGTGTTGTC	CAGGTCCCAG	ACCA	TGACAATGGT	CATGGCTGTC	CTCTCAAGCC
9901	GGGAGCGCCA	GCGCGTCTG	GGCCAGCATC	ACCCGGCACA	TCTCGCTGCT	GGCCCTCGATG
9961	ATCTCCATGA	GCTTGGCGTC	CGCGTACG	CGTCGACGA	CGTGTCC	TCTCGCGC
10021	GCCGACGCGA	GCACCTGTGC	GGCGGTGCG	GCCCCGGCGG	GGGCTGTT	GGCGCGACG
10081	TGCTTGGCCA	GGATGTCG	GGGCACCATC	TCGGGCGAGC	CCTCGTCCC	GTGGTGC
10141	GCGTACTCGC	ACACGCGGC	CGCGATCTGC	TCCGCGGTCC	ACAGGTCG	GATGTGCCG
10201	GCGACGAGTT	GGTGGTCG	GAGCGGCCGG	CCGAACTGCT	CCCGGGTCCG	GGCGTGGGCC
10261	ACCGCGGGCG	TGCGGCAGG	CCGAGGATC	CCGACCGAGC	CCCAGGC	CGACCTGCGC
10321	CCGTAGGGCGA	GTGACGCC	GACCAGCATC	GGCAGTGACG	CGCCGGAGGC	GGCCAGGACC
10381	GCGCCGGCCG	GCACACGAC	CTGGTCCAGG	TGCA	CGTGGCCGGC	GGCGCGGCAG
10441	CCGGACGGCT	TCGGGACGCG	CTCGACGCGT	ACGCGGGGG	TGTCGGCG	CACGACCA
10501	ACCGCACCGG	AACCATCTC	CTGGAGACCG	AAGACGACCA	GTTGGTC	GTAGGC
10561	GCAGTCGTCC	AGACCTGTG	GGCGTCGAC	ACAGGGTGT	CCCCGTCGAG	CCGAACCCG
10621	GTCCGGCATCG	CCGACAGATC	GCTGCCG	TGCGCTCAC	TGAAGCCGAC	GGCGCGAGT

10681	TTCCCGCTGG	TCAGCTCCTT	CAGGAAGGTC	GCCCCGCTGAC	CGGCGTCGCC	GAGCCGCTGC
10741	ACGGTCCACG	CGGCCATGCC	CTGCGACGTC	ATGACACTGC	GCAGCGAACT	GCAGAGGCTG
10801	CCGACGTGTG	CGGTGAACTC	GCCGTTCTCC	CGGCTGCCGA	GTCCCAGACC	GCCGTGCTCG
10861	GCCGCCACTT	CCGCAGAG	CAGGCCGTCG	GCGCCGAGCC	GGACGAGCAG	GTCGCGCGGC
5	10921	AGTCGCCGG	ACGTGCCCCA	CTCGGCCGGC	CGGTACCGA	CAAGGTCGGT
	10981	TCACGCTCAG	GCATGCCACGG	CCCGCAGCCG	GTGGACGAGT	GCGACCATGG
	11041	ACGGAAGTTC	GCGAGCTGGA	GGTCCGGG	GGCGATCGT	ACGTCGAACG
	11101	GTACACGACC	AGTCCATCG	CGAACAGCGA	CGTGAGGCCG	CCCTCCGCGA
	11161	GTCCACGGGC	CAGTCCGACC	TTGTCTTCGT	CTTGAGGAAC	GCGACCAACG
10	11221	GGGGTCGTCC	TTGACGGGTG	CGGTACATGAG	AACACCTTCT	CGTATTGTA
	11281	CCGGTCTTCC	GGCCGTGGT	TCCCTCGCGG	ACCTTGCCCA	GCAGCAGGTC
	11341	CTGCGCTCGT	CGCCGGTGC	TTTGTGCAGC	ACCCACAGCG	CGTCGACGAG
	11401	CCGATCAGGT	CCGCGGTGC	CAGCGGCCCG	GTCGGATGGC	CGAGGCACCC
	11461	GCGTCGACGT	CCTCGACGGA	CGCGGTGCC	TCCTGCACGA	TCCGCGCCGC
15	11521	ATCGGGTGG	GCAGCCGGCT	CGTGACGAAG	CCGGGCGCGT	CCCGGACGAC
	11581	CGCCGCAGCG	CCGCGAGCAG	GTCCCCGGCG	GCGGCCATGG	CTTTCTCAC
	11641	CCGCGGATCA	CCTCGACCGT	CGGGATCAGG	TACGACGGGT	TCATGAAGTG
	11701	AGGTCTCTGG	GCCGGGCCAC	GGAGTCGGCC	AGTTCGTCAA	CCGGGATCGA
	11761	GTGATGACCG	GGATACCGGG	CGCCGCTGCC	GAGACCGTGG	CGAGTACCTC
20	11821	TCGGCGTCCT	CGACGACGCC	CTCGATCAC	GCGGTGGCCG	TACCGATCGC
	11881	GACGTGGCCG	TCCCGACGAC	ACCGGGGTCG	GCCTCGGCCG	GCCCGGCCAC
	11941	GTCCGCACTT	CGGTGGCGAT	CCGCGCCCGC	GCCGCGCTAA	GGATCTCTC
	12001	ACGAGTGTCA	CCGGGACGCC	GTGGCGACG	GCGAGCGTGG	TGATGCCGGT
	12061	CCCGCGCCGA	GCACGATCAG	CTGGTGGTCC	ACGCTGTTTC	CTCCCTCCCG
25	12121	GCAGCGAGTA	CGGGTGCAGG	ACGTCTTCCG	GGGTCGACCC	GATCGCGTCC
	12181	GGCCGAGTTC	GTCGGCGAAG	CCGAGCAGCA	CGTCGAACGC	GATGTGGT
	12241	TGCCCTCTGA	GTCGAGGAGC	CTCAGGCTGT	CCCGGTGGTC	CCGCGCGGTG
	12301	CCGACAGGGC	CGCCAGCAG	GGGCCAGGCT	CGCGGTCCGG	CAGTTGCTGG
	12361	CGGCGCGGGC	CTGCCCCGGA	TTGTCACGCC	AGATGAACGC	GTCGTCGAGC
30	12421	GCAGTTCGGT	CTTGGCCGGC	TCGTCGGCC	CGATGGCGTT	CACATGCAGG
	12481	GGGGCTCGGC	GGGCAGCACC	GGCCCTTTC	CCGAGGGCAC	CGAGGGTACG
	12541	CATCCGCGGC	GGCGCGGGC	TCCGCCGGAT	CGGTACACCTT	GACCGGCACT
	12601	CGATGCGGTC	CGCGAACGAC	GCCGCGTGGC	CGGGGTCGGT	GTCGCTGACC
	12661	CGATGGGCAG	GACCCCTGCTG	AGCGCGTGC	CCTGGGTAC	CCGCTGTGCG
35	12721	TCAGCGTACG	CGTGGCGCTG	TCGGACCGGG	CCAGCAGCCG	GTCGCGACG
	12781	CGCCGGTCCG	CATCGGGTG	ATCACGCTG	CGTCGGCGAG	GGCGGTCA
	12841	CGTCGTCAG	CGCGACATC	GTGCCGACGA	TCGTCGGCAG	CCGGAAGCGC
	12901	GCGGACTGTA	CGAAACCGTC	TTCATGGTCA	CGCCGACACC	GGGGACCCGG
	12961	ACTCGATGAC	GCCGGGAATG	TCGCGCCCGC	GGACGAATCC	GGTACGCGC
40	13021	CGAACTCGCC	CGGGCCGAGC	CGGGCGAAC	CGTCGTCAG	CTCGCTGATC
	13081	TCATCACGTC	CGGGCCGATC	ACGGAGAGAA	TCCGCTTGAT	GTCACGTTGG
	13141	TGGTCTGCAT	GTGTACCTC	CCTTTCGTGG	CCGGAGCTGT	CTTGGTGGTG
	13201	CGGCTTCCGT	TCTCATCGCA	GCTCCCTGTC	GATGAGGTG	AAAATCTCGT
	13261	GTCCCGCGAC	AGCACGCCG	CCGGCGTGGT	CGGGCGGGC	TCCCGCCGCC
45	13321	CAGGGCGTCC	AGCCGGGTT	CGATCGCGTC	CGCCTGGCG	GCGCCCGGGT
	13381	AACGAGTGCT	TCCAGCCGGT	CGAGCTGCC	GAGCACACG	GTCACCGGGT
	13441	CAGCAGTTCA	CCGATGCGGT	CGCGAGTGC	GCGCGCGAC	GGGTAGTCGA
	13501	GGCGGACAGT	CGCAGACCGG	TCCGCTCGTT	GAGGGCGTTG	CGCAGCTGCA
	13561	CGAGTCCACA	CCGAGTTCCC	GGAACGCC	GTCCTCCGGG	ATGTCCTCCG
50	13621	GCCCAAGGAC	GCCGCTGCC	TCTGCGGAC	GAGGGCGAGC	AGGTGGTGG
	13681	CTCGTTGCG	GCGCTCCGGC	GGGCCGACGG	CTTGGGCCGG	CCACGCAGCA
	13741	CGGGGGCAGG	TCGCCCCGCC	CGGGGACGAC	ACTGCCCGT	CCGGTGTGGA
	13801	GTACATGCG	ATGCCCTGTT	CGCGGGTGA	CGCGCTCGCC	CCACCCCTTG
	13861	CCGGTCCGGC	TCGGTCAGGT	CCGGGGTCAG	GCCACTCGCC	TGGTCCCACA
55	13921	GATCGACAGC	CCTGGCAGCC	CTTGTGCACG	CCGGTGTTCG	GCGAGCGCGT
	13981	GTTCGCGGCC	GGCGTAGTTG	CCTGACCGGG	GGTGGCCAGC	ACACCGCCG
	14041	GACGACGAAT	GGGGCGAGGT	CGGTGTGCG	GGTGAGCCG	TGCAGGTG
	14101	GGCCTTGGGT	TTGAGGACGG	TGTCGATGCG	GTCGGGGGTG	AGGTTGTCGA
	14161	GTCGAGGGTT	CCGGCGGTG	GGAAGACGGC	GGTGAGGGT	TGAGGGATGT
60	14221	GGTGGCGAGT	TGGTGGGGT	CGCCGACGTC	GCAGGGGAGG	TGGGTGCC
	14281	GGGGGGTGGG	GTGCGGGAGA	GGAGGTAGGT	GTGGGGGTG	TTCAGGTG
	14341	GCCGGCGAGG	GTGCGGGAGC	CGCCGGTGT	GACGACGCC	CCCTCGGGGT
	14401	CGGGACCGTG	AGGACGATCT	TGCCGGTGTG	CTCGCCGCC	CTCATGGT
	14461	GCGGACCTGC	CGCATGTCGT	GCACCGTCAC	CGGCAGCGGG	TGCAGCACAC

14521	CAGGCCGAGC	AGCTCCGCGA	TGATCTCCTT	GAGCCGGTCG	GGCCCCGCGT	CCATCAGGTC	
14581	GAACGGTCGC	TGGACGGCGT	GCCGGATGTC	CGTCTCCCC	ATCTCGATGA	ACCGGCCACC	
14641	CGGCCGCGAGC	AGGCCGACGG	ACCGCTCGAG	GAGTCACCG	GTGAGCGAGT	TGAGCACGAC	
14701	GTCGACCGGC	GGGAACCGCGT	CGGCGAACGC	GGTGTGCGG	GAATGCCA	GATGCGCTCC	
5	14761	GTCCAGGTCC	ACCAGATGGC	GCTTCGCGGC	GCTGGTGGTC	GCGTACACCT	CCGCCCGCAG
14821	GTGCCGCGCG	ATCTGCCGG	CGCGCGAAC	GACACCGCCG	GTGGCCGCGT	GGATCAGGAC	
14881	CTTCTCGCCG	GGGCGCAGCC	CGCGGAGGT	GACCAGGCCG	TACCACGCCG	TCGCGAACGC	
14941	GGTCATCACG	GACGCCGCCT	GCGGGAACGT	CCAGCCGTCC	GGCATCCGGC	CGAGCATCCG	
10	15001	GTGGTCGGCG	ATGACCGTGG	GGCGAACGC	GGTGCACG	AGGCCGAAGA	CGCGGTCGCC
15061	CGGTGCCAGA	CCGGAGACGT	CGCGGCCGGT	CTCCAGGACG	ATGCCCGCGG	CCTCGCCGCC	
15121	GAGCACGCC	TGACCGGGGT	AGGTGCCGAG	CCGATCAGC	ACATCGCGA	AGTTGAGGCC	
15181	CGCCGCACGC	ACACCGATCC	GGACCTCGGC	CGGGGCGAGG	GGGCGCCGGG	GCTCCGCCGA	
15241	GTCGGCCGCG	GTGAGGCCGT	CGAGGGTGC	CGTCCCGC	GGCCGGATCA	GCCACGTGTC	
15	15301	GCTGTCCGGC	ACGGTGAGCG	GCTCCGGCAC	CCGGGTGAGG	GGGGCCGCCT	CGAACCGGCC
15361	GCCGCGCAGC	CGCAGACGCG	GCTCGCCAG	TGCGACGGCG	ATGCGCTGCT	GCTCGGGGGC	
15421	GAGCGTGACG	CCGGACTCGG	TCTCGACGTG	GACGAACCGG	CCGGGCTGCT	CGGCCTGGGC	
15481	GGCGCGCAGC	AGTCCGGCCG	CCGCGCCGGT	GGCGAGGCCG	GGGGTGGTGT	GCACGAGCAG	
15541	ATCCCCGCCG	GAGCCGGTCA	GGCGGGTCAG	CAGCCGGGTG	GTGAGCGCAC	GCGTCTCGGC	
15601	CACCGGGTCC	TCGCCATCAG	GGCAGGGCAA	CGTGTGACG	TCCACGTGCG	TCGCGGGGAC	
20	15661	ATCCCTGGGT	GCGGCCACCT	CGATCCAGGT	GAGACGCATC	AGGCCGGTGC	CGACGGGTGG
15721	GGACAGCGGG	CGGGTGCAGG	CCGTCGGGAT	CTCGGGCACG	AGTTGGCCGG	CGGAGTCGGC	
15781	GACCGCGAGA	CTCAGCTCGT	CCCGTCACG	AGTGTACAGC	GCTCGGAGCA	TGGCCGAGCC	
15841	CGTGGCGACG	AACCGGGCCC	CCTTCAGGGC	GAACGGCAGA	CCCGCAGCGC	TGTGTCAGGG	
25	15901	CGTGGTGAGG	GCGACGGCGT	GCAGGGCCGC	GTCGAGCAGC	GCCGGATGCA	CACCGAAACC
15961	GTCCGCCTCG	GCGGCCCTGC	CGTCGGGCAG	CGCCACCTCG	GCATACACGG	TGTACCCATC	
16021	ACCGCAGGCA	GCCCCGAAAC	CGTGGAACGC	CGACCCGTAC	TCATAACCGG	CATCCCGCAG	
16081	TTCGTATAG	AACCCCAGA	CGTCGACGGC	CACGGCCGGT	ACCGGGCGCC	ACTGGAGAA	
16141	CGGCTCCACA	CCGACAAACAC	CGGGGGTGTG	GGGGGTGTG	GGGGTCAGGG	TGCCGCTGGC	
30	16201	GTGCCGGGTC	CAGCTGCCCG	TGCCCTCGGT	ACGCGCTGTT	ACGGTCACCG	GCCGGCGTCC
16261	GGCCTCATCA	GCCCCCTCCA	CGGTACCGA	CACATCACC	GCTCGGGTCA	CCGGCACCA	
16321	AAGGGGGGAT	TCGATGACCA	GCTCGTCCAC	TATCCCGCAA	CCGGTCTCGT	CACCGGCCCC	
16381	GATGACCAGC	TCCACAAACG	CCGTACCCGG	CAGCAGGACC	GTGCCCCGCA	CCGCGTGTAC	
16441	AGCCAGCCAG	GGGTGAGTGC	GCAATGAGAT	CCGGCCAGTG	AGAACAAACAC	CACCATCGTC	
35	16501	GGGGGGCAGC	GCTGTGACAG	CGGCCAGCAT	CGGATGCGCC	GCACCCGTCA	ACCCGCCGC
16561	CGACAGATCG	GTGGCACCGG	CCGCCTCCAG	CCAGTACCGC	CTGTGCTCGA	ACGCGTACGT	
16621	GGGCAGATCC	AGCAGCCGTC	CCGGCACCGG	TTCGACCACC	GTGTCCTCAGT	CCACTGCCGT	
16681	GCCCAAGGGTC	CACGCCCTGC	CCAACGCCGT	CAGCCACCGC	TCCCAGCCGC	CGTACCCGGT	
16741	CCGCAACGAC	GCCACCGTGT	GAGCCTGCTC	CATGCCCGC	AGCAGCACCG	GATGGGCACT	
40	16801	GCACCTCACG	AACACCGACC	CATCCAGCTC	CGCCACCGCC	GGTCCAACG	CCACCGGACG
16861	ACCGCAGATT	CGGTACCGAT	ACCCCTCATC	CACCGGCTCC	GTCACCCAGG	CGCTGTCAC	
16921	GGTCGACCAAC	CACGCCACCG	ACGCGGCCTT	CCCTGCCACC	CCCTCCAGTA	CCTTGGCCAG	
16981	TTCATCCTCG	ATGGCTTCCA	CGTGGGGCGT	GTGGGAGGCG	TAGTCGACCG	CGATACGACG	
17041	CACCCGCACG	CCTTCGGCCT	CATACCGCGC	CACCACTCC	TCCACCGCCG	ACGGGCCCC	
45	17101	CGCCACCACC	GTGAAAGCGG	GGCCGTTACG	CGCCCGATC	CACACACCT	CGACCAGACC
17161	GACCTCACCG	GCCGGCAACG	CCACCGAACG	CATCGCTCCC	CGCCCCGGCCA	GTCGCGCCGC	
17221	GATGACCTGA	CTGCGCAATG	CCACCACCGC	GGCGCGTCC	TCGAGGCTGA	GGGCTCCGGC	
17281	CACGCACGCC	GCCCGATCT	CGCCCTGGGA	GTGTCCGATC	ACCGCGTCCG	GCACGACCCC	
17341	ATGCGCCTGC	CACAGCGCG	CCAGGCTCAC	CGCGACCGCC	CAGCTGGCCG	GCTGGACCAC	
50	17401	CTCCACCCGC	TCCGCCACAT	CCGGCCCGC	CAACATCTCC	CGCACATCCC	AGCCCGTGTG
17461	CGGCAGCAAC	GCCTGAGCGC	ACTCCTCCAT	ACGCGGGCG	ACACCGCGG	AGTGGGACAT	
17521	GAGTTCCACG	CCCATGCCGA	CCCACTGGGC	GCCCTGGCCG	GGGAAGACGA	ACACCGTACG	
17581	CGCCTGGTCC	ACCGCCACAC	CCGTACCCCG	GGCATCGCCC	AGCAGCACCG	CACGGTGAAC	
17641	GAAGACAGCA	CGCTCCCGA	CCAACCCCTG	CGCGACCGCG	GCCACATCCA	CACCAACCCC	
55	17701	GCGCAGATAAC	CCCTCCAGCC	GCTCCACCTG	CCCCCGCAGA	CTCACCTCAC	CACGAGCGA
17761	CACGGGCAAC	GGCACCAAC	CGTCAACAAAC	CGACTCCCCA	CGCGACGGCC	CAGGAACACC	
17821	CTCAAGGATC	ACGTGCGCGT	TCGTACCGCT	CACCCCGAAC	GACGACACAC	CCGCATCGGG	
17881	TGCCCCATCC	GACTCGGGCC	ACGGCCTCGC	CTCGGTGAGC	AGCTCCACCG	CACCGGCCGA	
17941	CCAGTCCACA	TGCGACGAGC	GCTCGTCCAC	ATGCAGCGTC	TTCGGCGCGA	TCCCCTACCG	
60	18001	CATGCCCATG	ACCATCTTGA	TCACACCGGC	GACACCGCC	GGCGCCTCGG	CATGACCGAT
18061	GTTCGACTTC	AACGAACCCA	GCAGCAGCGG	AACCTCACGC	TCCCTGCCCGT	ACGTCGCCAG	
18121	AATGGCCTGC	GCCTCGATGG	GATCGCCCG	CGTCGCCCC	GTCCCCGTGCG	CCTCCACCC	
18181	GTCCACATCG	GCGGCGCGA	GTCCGGCGTT	CACCAACGCC	TGCTGGATGA	CACGCTGCTG	
18241	GGACGGGCCG	TTGGGGCGG	ACAGCCCGTT	GGAGGCACCG	TCCTGGTTCA	CCGGCACCC	
18301	GCGGACGACCC	GCGAGAACGG	TGTGTCCGTT	GCGCTCGCG	TCGGAGAGCC	GCTCCAGCAC	

18361	AAGAACGCCG	GCGCCCTCCG	CCCAGCCGGT	GCCGTTGGCG	GCGTCCGCCA	ACGCGCGGCA	
18421	GCGGCCGTCG	GGGGAGAGTC	CGCCCTGCTG	CTGGATTCC	ACGAACCCGG	TGGGGTTCG	
18481	CATGACGGTG	ACACCGCCGA	CCAGCGCCAG	CGAGCACTCC	CCGTGGCGCA	GTGCGTCCC	
18541	GGCTGGTGC	AGCGCGACCA	GCGACGACGA	GCACGCCGTG	TCCACCGTGA	ACGCCGGTCC	
5	18601	CTGGAGCCCA	TAGAAGTACG	AGATCCGGCC	GGTGAGCAGC	CTGGGCTGCA	TGCCGATCGA
18661	GCCGAACCCG	TCCAGGTCCG	CGCCGACGCC	GTACCCGTAC	GAGAAGGCGC	CCATGAACAC	
18721	GCCGGTGTG	CTGCCGCGCA	GTGTGCCCGG	CACGATGCC	GCGCTCTCGA	ACGCCTCCC	
18781	TGTCGTTTCC	AGCAGGATCC	GCTGCTGGGG	GTCCATGGCC	CGTGCCTCAC	GGGGGCTGAT	
18841	GCCGAAGAAC	GCGGCATCGA	AGCCGGCGGC	GTGGAGAGG	AAGCCGCCG	GGTCCGTGTC	
10	18901	CGATCCGCCG	GTGAGGCCGG	ACGGGTCCCA	GCCACGGTCG	GCCGGGAAGC	CGGTGACCGC
18961	GTCGCCGCCA	CTGTCCACCA	TGCGCCACAG	GTCGTCGGGC	GAGGTGACGC	CGCCCGGCAG	
19021	TCGGCAGGCC	ATGCCACGA	TGCCAGCGG	TTCGTCACGG	GTCGCGGCCG	CTGTGGGAAC	
19081	AGCGACCGGT	GCGGCACCAC	CGACCAAGAGC	CTCGTCCAAC	CGCGACGCCA	TGGCCCGCGG	
19141	CGTCGGGTAG	TCGAAGACAA	GCGTGGCGGG	CAGTCGGACA	CCGGTCGCCG	CGCGAGTCG	
15	19201	GTTCCGCAGT	TCGACGGCGG	TCAGCGAGTC	GATACCCAGT	TCCTTGAAGG	CCGCGTCCGC
19261	GGACACGTCC	GCGCGTCCG	CGTGGCCGAG	CACCGCCGCC	GCGTTGTCG	GGACCAAGTGC	
19321	CAGCAGCGCG	GTGTCCCGCT	CAGCGCCGGA	CATGGTCCG	AGCCGGTCGG	CGAGCGGAAC	
19381	GGCGGTGGCC	GCGCCGGG	GCGATACGGC	GCGGCCAGA	TCGGCGAAAA	CGGGCGATGT	
20	19441	GTGCGCGGTG	AGGTCCATCG	TGCGCCAC	GGCGAACGCG	GTGCCGGTTC	CGGCCGGCGC
19501	TTCCAGCAGG	CGCATGCCA	CACCGGCCGA	CATGGGCGG	AAACCGCCG	GGCGGACACG	
19561	GGTGCCTTGT	GTGCCGCTCA	TGCTGCCGGT	GAGTCCGCTG	TCATCGGCC	AGAGGCCCA	
19621	GGCAGCGAC	AGCGCGGGCA	GTCCCTCGGC	ATGGCCGAGC	GTCGCGAGTC	CGTCGAGGAA	
19681	CCCGTTCGCC	GCGAGTAGT	TGCCCTGGC	CGGGCCGCC	ATGATGCCG	CGACGGACGA	
25	19741	GTAGAGGACG	AACGAGCGA	GGTCCGCGTC	CCGGGTCA	TGTCGAGGT	GCCAGGCGCC
19801	GTCGGTTTG	GGGCGCACTG	TGTTGGCGAG	CCGCTCCGGG	GTGAGTGC	TGGTACGCC	
19861	GTCGTCGAGC	ACGGCTGCCG	TGTTGAAGAC	CCGCGTGAGC	GCCCTGCCG	CGGCGGCGAG	
19921	CGCCGGCGCG	AGCTGGTCCC	GGTGGCGAC	GTCACAGCG	ATGTTGACAC	CGGGAGTGT	
19981	CGCCGGCGGT	TCGCTGCCG	ACAGCAACAG	GAGGTGGCGG	GCGCCATGCT	CGGCCGACGAG	
30	20041	ATGCCGGCG	AGGAGACCTG	CCAGCACACC	CGAGCCGCCG	GTGATGACCA	CCGTGCCGTC
20101	CGGGTCGAGC	AGCGGTTCGG	GCGTTCCG	GGCGGCCGTG	CGGGTGAACC	GCGGCCGCTC	
20161	GTACCGGCCG	TCGGTGACGC	GGACGTACGG	CTCGGCCAGT	GTGTCGGCG	CGGCCAGCCC	
20221	CTCGATGGGG	GTGTCGGTGC	CGGTCTCCAC	CAGCACGAAC	CGGCCCCGGG	GCTCGGCTG	
20281	GGCGGACCGG	ACGAGGCCGG	CGACCGCTCC	TCCGACCGGT	CCCGCGTCGA	TCCGGACGAC	
20341	GAGGGTGGTC	TCCGAGGGC	CGTCCCTGGC	GATCACCCGG	TGCACTCGC	CGAGCACGAA	
35	20401	CTCGGTGAGC	CGGTACCGTCT	CGTCGAGGAC	ATCCGCC	GGTCCGGGA	CGCGGAGAC
20461	GATGTGGACC	GCGTCCGCAG	GACCGGGCCC	GGGAGTGGG	AGCTCGGT	AGGAGAGGCC	
20521	GTACAAGGAG	TTCCGTACGA	CGGCGCGTC	GCCGTCGACG	TTCACCGGT	GCGCGGT	
20581	CGCGCGACG	GTCACCAACCG	GTTGGCCGAC	CGGGTCCGTC	GCATGCACGG	CAGCGCCGTC	
20641	CGGGCCCTGA	GTGATCGTGA	CGCGCAGCGT	GGTGGCCCG	GTGTCGTGGA	ACCGCACGCC	
40	20701	GCTCCACGAG	AACGGCAGCC	GCACCTCCG	TTCCCTGTTCC	GCGAGCAGCG	GCAGGCAGGT
20761	GACGTGCAAG	GCGCGTCGA	ACAGCGCCG	GTGGACGCCA	TAGTGC	GGGCGTCEGC	
20821	CTGTTCCCG	GCGATCTCCA	CCTCGCGTA	CAGGGTTTCG	CCGTCGCGCC	AGGCGGT	
20881	CAGTCCCTGG	AACGCTGGC	CGTAGCTGTA	GCCGGTCTCG	GCCAGCCGCT	CGTAGAACGC	
45	20941	GCTCACGTCG	ACCGTCGCG	CGCCCGCGG	CGGCCACGCG	GGCGGCGGG	CCGCGCGAC
21001	GCTTCCGGCC	CGGCCGAGGG	TGCCGCTGGC	GTGCCGGGT	CAGCTGTCCG	TGCCCTCGGT	
21061	ACGCGCGTGG	ACGGTCACTC	GCCGCCGTCC	GGCCTCATCG	GCCCCCTTCGA	CGGTCAACCGA	
21121	CACATCCACC	GCGCCGGTCA	CCGGCACCAAC	GAGCGGGGT	TGCACTGCCA	GTTCATCCAC	
21181	CACCCCGCAA	CCGGTCTCGT	CACCGGCCCG	GATGACCGAC	TCCACAAACG	CCGTACCCGG	
50	21241	CAGCAGAAC	GTGCCCCGCA	CCCGCGTATC	AGCCACGCC	GGATGCGTAC	GCAACGAGAT
21301	CCGCCAGTG	AGAACAAACAC	CACCAACGTC	GTCGGCGGGC	AGTGCTGTGA	CGGCGGCCAG	
21361	CATCGGATGC	GCGCCCGCG	TCAGCCCGG	CGCGGACAGA	TGGTGGCAC	CGGCCGCTC	
21421	CAGCCAGTAC	CGCCCTGTGCT	CGAACCGCTA	GGTGGCGAGA	TGAGGACGCC	GTCCCCGGAC	
21481	CGGTTGACCC	ACCGTGTCCC	AGTCCACTGC	CGTGGCCAGG	GTCCACGCC	GCGCCAACGC	
21541	CGTCAGCCAC	CGCTCCAGC	CGCCGTCACC	GGTCCCGAAC	GACGCCACCG	TGTGAGCCTG	
55	21601	TTCCATCGCC	GCGAGCAGCA	CCGGATGGG	GCTGCAC	ACGAACACGG	ACCCGTCAG
21661	CTCCGCCACC	GCGCGTCCA	GGCGGACGGG	GGCAGCGAGG	TTCCGGTAC	AGTAGGCC	
21721	ATCCACCGGC	TCGGTCACCC	AGGCCTGTC	CACCGTGGAC	CACCAAGGCCA	CCGACCCGGT	
21781	CCCGCCGGAA	ATCCCCCTCA	GTACCTCGGC	CAACTCGTC	TCGATGGCTT	CCACGTGGGG	
21841	CGTGTGGGAG	GCGTAGTCGA	CCGCAGATC	GCGCACTCGC	ACGCTTCGG	CCTCGTACCG	
60	21901	CGTCACCACT	TCTTCCACCG	CGGACGGGT	CCCCGCCACC	ACAGTCGAAG	ACGGGCCGTT
21961	ACGCGCCGCG	ATCCACACGC	CCTCGACCA	GTCCACCTCA	CCGGCCGGCA	ACGCCACCGA	
22021	AGCCATCGCC	CCCCGCCGG	CCAGCCGCC	GGCGATCACC	TGGCTGCGCA	AGGCCACCCAC	
22081	CGGGCGGGCG	TCCTCAAGGC	TGAGGGCTCC	GGCCACACAC	GCGCCCGCGA	TCTCGCCCTG	
22141	GGAGTGTCCG	ACCACCGCGT	CCGGCACGAC	CCCATGCGCC	TGCCACAGCG	CGGCCAGGCT	

22201	CACCGCGACC	GCCCAGCTGG	CCGGCTGGAC	CACCTCCACC	CGCTCCGCCA	CATCCGGCCG	
22261	CGCCAACATC	TCCCACAT	CCCAGCCCGT	GTGCAGCAAC	AACGCCCGC	CACACTCCTC	
22321	CATACTGAGCC	GCAGAACACCG	CAGAACACCGC	CATCAACTCC	ACACCCATGC	CCACCCACTG	
22381	AGCACCCCTGC	CCGGGAAAGA	CGAACACCCGT	ACGCAGCTGA	TCCACGCCA	CACCCATCAC	
5	22441	CGGGGCATCG	CCCAACAACA	CCGCACGGTG	ACCGAAGACA	GCACGCTCAC	GCACCAACCC
22501	CTGCGCGACC	GCAGCACAT	CCACACCACC	CCCGCGCAGA	TACCCCTCCA	GCCGCTCCAC	
22561	CTGCCCCCGC	AGACTCACCT	CACTCCGAGC	CGACACCAGC	AACGGCACCA	ACCCATCGAC	
22621	AGCCGACTCC	CCACGCGACG	GCCCAGGAAC	ACCCCTCAAGG	'ATCACGTGCG	CGTTCTGAC	
22681	GCTCACCCCCG	AAAGCGGAGA	CACCGGCCCG	CGCGGACGT	CCCAGCTCGG	GCCACGCCCG	
10	22741	CGCTCTGGTG	AGCAGTCTCA	CCGGCGCCCTC	GGTCCAGTCC	ACATGCGACG	ACGGCTCGTC
22801	CACATGCAGC	GTCTTCGGCG	CGATGCCATA	CCGCATCGCC	ATGACCATCT	TGATGACACC	
22861	GGGACACCCC	GCAGCCGCCT	GGCATGACC	GATGTTGAC	TTCAACGAAC	CCAGCAGCAG	
22921	CGAACCTCA	CGCTCTCGC	CGTACGTGCG	CAGAACATCGC	TGCGCCTCGA	TGGGATCGCC	
22981	CAGCGTCGTC	CCCGTCCCCT	GGCGCTCCAC	CACGTCACG	TGCGCGGGGG	CGAGCCCGC	
15	23041	CTTGTGGAGG	GCCTGGCGGA	TGACGCGCTG	CTGGGAGGGG	CGTGTGGGTG	CGGAGATGCC
23101	GTTCGAGGCG	CCGTCTGGT	TGACGGCGGA	GGAGCGGACG	ACCGCGAGGA	CGGTGTGTCC	
23161	GTTCGCGCTCG	GGCTGGAGA	GCTTTTCGAC	GACGAGGACG	CGGGCCCCCT	CGGGAAACCC	
23221	GGTCCCGTCC	GGCCGCTCAG	CGAACGCCCT	GCACCGTCCG	TCCGGCGCGA	CGCCGCCCTG	
20	23281	CCGGGAGAAC	TCCACGAAGG	TCTGTGGTGA	TGCCATCACT	GTGACACCAC	CGACAGCGC
23341	CAGCGAGCAC	TCCCCGGTCC	GCAGCGCTG	CCCGGGCTGG	TGACGCGCGA	CCAGCGACGA	
23401	CGAACACCGCC	GTGTCGACCG	TGACCGCCGG	ACCCCTCATG	CCGAAGAAAGT	ACGACAGCCG	
23461	TCCGGCGAGC	ACCGCGGGCT	GTGTCGCTGA	GGCGCCGAAT	CCGCCCCAGGT	CCGCCCCCGT	
23521	GCCGTAGCCG	TAGTAGAAC	CGCCGACGAA	GACGCCGGTG	TGCGTCCCGC	GCAGGGTGTG	
23581	CGGCACGATG	CCGGCGTGT	CGAGCGCCTC	CCAGGGGATT	TGAGGAGGAA	TCCGCTGCTG	
25	23641	CGGGTCGAGT	GGGGTGGCCT	CGCGCGGACT	GATGCCGAAG	ACCGGGCAT	CGAAGTCGGC
23701	GGGCCCCCGC	AGTGCGCCGG	CCCGCCCGGT	GGCGGACTCG	GGGGCGGCGT	GCAGCGCGC	
23761	CACGTCCCAG	CCGCGGTGG	TGGGAAAGTC	GCCGATCGCG	TGCGGGCCGT	CCGCGACGAG	
23821	CTGCCACAGC	TCTTCCGGT	AGGTGACGCC	GCCCGGCAGT	CGGCAGGCCA	TGCCGACGAC	
23881	GGCGAGCGGC	TCGTTCGCCG	CGCGCGCGAG	CGCGGTGTT	TCCCGGCCGA	GCTGCGCGTT	
30	23941	GTCCTTGACC	GACGTCCCA	GGCGCTCGAT	CAGGTCGTT	TGCGCCATCG	CCTCATCCCT
24001	TCAGCACGTG	CGCGATGAGC	GGCTCTCGCT	CCATGTCGTC	GAACAGTTG	TCGTCGGCT	
24061	CCGGGTCGT	GGTGCTCGC	GGTGCTGTG	CCGGTGGTTC	ACCGCCGTCC	GGGGTCCCCT	
24121	TGTCGTCGG	GGTCCCCTG	ACGTCCGGGG	CCAGGAGGGT	CAGCAGATGA	CGGGTGAGCG	
24181	CGCCGGCGGC	GGGATAGTCG	AAGACGAGCG	TGGCCGGCAG	CGGAATGCCG	AGGGCCTCGG	
35	24241	AGACCCGGTT	GGCGAGGCCG	AGCGCGGTGA	GCGAGTCGAC	CCCGAGGTCC	TTGAACGCCG
24301	TGGTGGCCGT	GACCGCCGCC	GGCTCGGTGT	GGCCCAGCAG	GTTGGCGGGC	GTGTCGCGGA	
24361	CGACGCCGAG	CAGCACCTGT	TCCCCTTCCT	TGTGGGGCAG	GTCCGGCAGG	CGTTCCAGCA	
24421	GGGAGCCGCC	GTCGGTCGCG	GAGCGCCGGG	TGGGGCGCTG	GATCGGTGCG	CACAGCGGTG	
40	24481	ACGGGTCGCC	GGGCCCCGGT	GGGGCGGTGCG	CCACGACCAC	GGCTTCCCCG	GTGGCGCACG
24541	CGCGTTCGAG	GAGGTGGTGC	AGCCGGTCCG	CCGCGCGGT	GAACGCCACG	CCCGGCAGGC	
24601	CTTGTGCCCG	CGCGAGGTG	GCCAGGGCCT	GGAGCGGTCC	GGCCGCTCTG	CCGGACGGAA	
24661	CGGGGAGAAC	GAACGCGTC	AGGTGCGAGGT	CGCGGGTCAG	GCGGTGCAAGT	TCCCAGGCCG	
24721	ACTCGCCGGT	GGCGTCGCG	TGACGACCG	CGGTACCCG	GTTTCCGGC	ACTGTGCCG	
45	24781	GCTCGTACCG	GATCACTTCG	GGCCGCTGTC	CGCCGAGGT	TCCGGCAGT	TCCCTCGAAC
24841	CGCCCGCGAG	GAGGACGGT	TGCCCTACCG	AGGCCGCCGC	CCTGGTGGGC	CGGGGGGGGA	
24901	CGAGGGCGGG	CGCTTCGAGG	CGCCCGTCGG	CCAGGGCAG	GTGCGGTT	TCGAGGCGGG	
24961	AGAGGGCGGG	GGCGCGCGGG	GGGGTGACCC	TGTCGGTGGT	CTCCACGAGC	ACGAGCCGGC	
50	25021	CCGGTTCCGC	GGTGTGAGC	AGTGCAGCGA	CGGCACCGGC	GACGGGCCCG	GCCTCGCGG
25081	ACACCACCA	CGTGGCGCCG	GCGGTCTCTG	GGTCGTCAG	TGCGGTACGG	ACCTCGTCGG	
25141	GACCGGATAC	CGGGACGAGC	ATGACGTCGG	GGTCGTCGTC	GTGCGCGAGG	TGGGTGATCC	
25201	GGCGGGCGT	GGTCCCCGGT	GGCCCGGGGG	CCCGGACGCC	GTTCCAGGGT	CGCCGGAAAC	
25261	GGCGCACGTC	CCCGTCCGGG	CCCGTCGTTG	CGGGGGGCCG	GTTGATGAGC	GAGCCG..TCT	
25321	GAGCCACCGG	CGTCCCCAGT	TGTCGGCGA	GGTGCACGCG	GGCGCCGCC	TCGCCCTCGC	
55	25381	CGTGGACGAA	GGTGACCGC	AGTTTCGTTG	CGCCGCTGGT	GTGGACACGG	ACGCCGGTGA
25441	ACCGGAACGG	CAACCGTAC	CCCGCGTTCT	CGCGGCCGC	GCCGATGCTG	CCCGCTTGC	
25501	GCGCGGTGAC	GAGCAGCGCC	GGGTGCACTG	TGAGCGGGC	GGCGTCCCTG	GCGAGGGCGC	
25561	CGTCGAGGGC	GAATTCGGCG	CAGACGGTGT	CTCCGTGGT	CAACGCGGCC	GACATGCCGC	
25621	GGAACTCGGG	GCCGAACCTG	TATCCCGCTG	CGTCGAGTCG	CTGGTAGAAG	GCCCGCACGT	
60	25681	CGACCGGGTTC	CGCGTTCGCG	GGCGGCCAGG	GCCCCGGCGT	GGTGGCCGGT	TCGGTGGTGG
25741	CGATGCCGGC	GAAGCCGGAG	GGGTGGCGGG	TCCATGTCG	GTGCGCGTCC	GTCCGGCGGT	
25801	GGACGCGCAC	GGCACGGCGT	CCGGTGTGCG	CGGGCGCGGC	GACGGTCACG	CGCACCTGGA	
25861	GGGCGCCGGT	GGCGGGCAGG	ACCAGCGGGTG	TCTCGACGAC	CAGTCGTCG	AGCAGGTCGC	
25921	AGCCTGCC	GTCGGCGCCG	CGTCCGGCCA	ATTCCAGGAA	GGCGGGTCCG	GGCAGCAGTA	
25981	GGCGGCCGTC	GACGGAGTGA	CCGGCCAGCC	ATGGGTGGGT	GGCCAGCGAG	AACCGGCCGG	

26041 TGAGCAGCAC CTCGTCGGAG TCGGGGAGCG CCACCGACGC GGCGAGCAGC GGGTGGTCGA
 26101 CGCGCTCGAG TCCGAGGCCG GAAGCGTCCG TGCCGCCGC GGTCTCGATC CAGTAGCGCT
 26161 CATGGTGGAA GGCATATGTG GGCAGGTCGT GTGCCGTGCG CGTCGCGGGG ACGACCGCCG
 26221 CCCAGTCGAC GGGCACGCCG GTTGTGTGCG CCTCGGCCAG CGCGGTGAGC AGCCGGTGA
 5 26281 CTCCCCCGCC GCGGCGGAGC GTGGCGACGG TCGCGCCGTC GATCGCGGGC AGCAGCACGG
 26341 GGTGCGCGCT GACCTCGACG AACACGGTGT CACCCGGCTC GCGGGCAGCG GTCACGGCCG
 26401 TGGCGAAGCC TACGGGGTGG CGCATGTTGC GGAACCAAGTA CTCGTCGTGAGC AGCGCGCGT
 26461 CGATCCAGCG TTGTCGCGCG GTGGAGAAC ACGGGATCTC GGGCGTGCAG GAGGTGGTGT
 26521 CCGCGACGAT CCGCTGGAGT TCGTCGTACA GCGGGTGCAG GAACGGGGTG TGGTCGGGC
 10 26581 AGTCGACGGC GATCGGGCGC ACCCAGACGC CGCGGGCCTC GTAGTCGGCG ATCAGCGTTT
 26641 CGACGGCGTC CGGGCGCCCG GCGACGGTCG TGGTGGTGGC GCGCTTGCAG CCCCGACCC
 26701 AGACGCCGTC GATCCGGCGC GCATCCGCCT CGACGTGCGC GGCCGGGAGC GCGACCGAGC
 26761 CCATCGCGCC GCGTCCGGCG AGTCGCGCA GGAGCAGGAG AACGCTGCGC AGCGCGACGA
 26821 GCGGGGCACC GTCCTCCAGG GTGAGCGCTC CGCGACACA GC CGCGCGGGC ATCTCGJCC
 15 26881 GGGAGTGTCC GATGACGGCG TCCGGGCGTA CGCCCGCGGC CTCCCACACG GCGGCCAGCG
 26941 ACACCATGAC GGCCCAAGCAG ACAGGGGTGCA CGACGTGAC GCGCGGGGTC ACCTCCGGGT
 27001 CGTCGAGCAT GCGGATGGGG TCCCAGCCCG TGTGCGGGAT CAGCGCGTCG GCGCATTGGC
 27061 GCATCCTGGC GCGAACACCC GGGGAGGCCG CCATCAGTTC GACGCCCATG CGCGGCCACT
 27121 GCGGTCTTC TCGGGGAAAG ACAGAAGACGG TGCGCGGCTC GTGAGCGGCC GTGCCGGTGA
 20 27181 CGACGTCGTC GTCGAGCAGC ACAGGCGCGGT GCGGGAAACGT CGTACGCCCTG GCGAGCAGGC
 27241 CCGCGGCGAT GGC CGCGGGG TCGTGGCGG GACGGCGGC GAGGTGCTCG CGGAGTCGGC
 27301 GGACCTGGCC GTCGAGGGG G TGGCGGGTCC GCGCCGAGAC GGGCAGTGGT GTGAGCGGCC
 27361 TGGCGATCAG CGGCTCACCG GCGCTCGAGG CCGACGGCTC CT CGGCCCGGC GGCTCCCCGG
 27421 CCGGGTGGGC TTCCAGCAGG ACAGTGGCGT TGGTGGCGCT GACGCCGAAG GAGGACACAC
 25 27481 CGGCGCGCCG CGGGCGGTG GTCCTCGGGC AGGGCCGGG ACATGGTGAGG AGTTCGACGG
 27541 CGCCGGCGT CGAGTCGACG TGCGAGGACG CGGTGTCCAC GTGCAGGGTG CGCGGCAGGG
 27601 TGCGTGTGCCG CATGGCGAGG ACCATCTTGA TGACACCGGC GACACCCGGC CGGGCCTGAG
 27661 TGTGGCCGAT GTTGGACTTC AGCGAGCCCA GCAGCACCGG GGTGTCGCGC CCCTGGCCGT
 30 27721 AGGTGGCCAG CACCGCTGT GCTCGATGG GATCGCCAG CCTGGTGCAG GAGGAGAGCC
 27781 CCTCCACGGC GTCCACGTCC GCGGGGTTGA GCGCGCGTT GGCCAGGGCC TGCGGATCA
 27841 CCCGCTCCTG CGAGGGCCCG TTGCGCGCCG ACAACCCGTG GGAAGCACCG TCCTGGTTGA
 27901 CGGCCGAACC CCGGACAACC GCCAGCACAC GGTGGCGTT GCGCTGGCA TCGGAGAGCC
 27961 TCTCGACGAT CAGCACACCG GACCCCTCGG CGAAACCGGT GCGTCAGCC GCATCCJCGA
 28021 ACGCTTGCA GCGCGCGTC GCGCGAGAC CCCGCTGCTG GGAGAACTCG ACGAACCGG
 35 28081 ACGCGAGGC CATCACCGT ACAGCCCGA CGAGGGCGAG CGAGCATTG CGCGAGCGCA
 28141 GTGACTGCCG GGCCTGGTGC AGCGCCACCA GCGAGCACGA ACACGCCGTG TCGACCGTGA
 28201 CGGCCGGACC CTCCAGACCG TAGAAGTACG ACAGCCGACC GGACAGCACA CTGGTCTGGG
 28261 TGCGGTGCGC GCGGAAACCG CCCAGGTGCG TGCGAGTCC GTACCCGTG GAGAAGGC
 28321 CCATGAACAC GCGGTGTCG CTTCGCGCA GCGACTCCGG GAGGATCCCC CGGTGTTCCA
 40 28381 GCGCCTCCCA CGAGGTCTCC AGGACCAGAC GCTGCTGCGG GTCCATCGCC AGCGCCTCAC
 28441 CGGACTGTAT CCCGAAGAAC GCGCGTCGA AGTCCGCCAC CCCGGCGAGG AAGCCACCAT
 28501 GACCGACGGT CGACGTGCCC GGATGATCCG GATCGGGATC GTACAGCCCG TCCACGTCCC
 28561 AACACACGGTC CGTCGAAAC GCGGTGATCC CGTCACCAACC CGACTCCAGC AGCCGCCACA
 28621 AGTCCTCCGG CGACCGGACCC ACACCCGGCA GCGGGCAGGC CATCCCCACG ATCGCCAACG
 45 28681 GCTCGTCTG CGGGACGGCC CGGGTCGTGG TGCGGGTCCG CGATGCCGTG CGGCCGGACA
 28741 CGGCCGGGT GAGCTTCGCC GCGACGGCGC GCGCGTCGG GAAGTCGAAG ACCGCCGGTGG
 28801 CGGGCAGCGG TACGCCGTG CCTCGGGTGA AGGCCTGGC GAGCCGGATC GCGATGAGCG
 28861 AGTCGACGCC GAGTTCTTG AACGTGGCGG TCGCCTCGAC CGCTGGCGCA CGGTGTTGCC
 28921 CGAGTACGGC CGGGTGCAC TCGCGGACGA CGCGAGCAGC GTCTTTTCG CGGTCCGGCG
 50 28981 CGGAGAGCCG CGCGATCCGG TCGCGAGGG TGGTGGCGCC GGCGCGCCCG CGCCGGCGCT
 29041 CCCGGCGCGG TGCGCGCAGC AGGGCGAGC TGCGGAGGC GGCGGGGTGCG GCGGCGACCA
 29101 CGCCCCGGTC CGAGGACCCG AACCCCGGT CGAACACCGT CAGTCGGCCCT TCGGGCJTC
 29161 CGCGCGTCAC GCGTCCGG CGCATGCGGG CGCGGGTGCC GACCGTCAGC CCGCTCTCCG
 29221 GTTCCACAG GCGCCCGAGCC AGGGACAAAG CGGGCAGTCC GGCTGCCCGG CGCTGTTCG
 55 29281 CCAGCGCGTC GAGGAACCGC TTCGCGGGCC CGTAGTTGCC CTGTCCGGGG CTGCCGAGCA
 29341 CACCGCGGGC CGACGAGTAG AGGACGAACG CGGCCAGTTC CGTGTCTGG GTGAGTTCGT
 29401 GCAGGTGCCA CGCGCGTCC ACCTTCGGGC GCAGCACCGT CTCGAGCCGG TCGGGGGTGA
 29461 CGCGGGTGA GACCGCGTCG TCGAGGACGG CGCGGGTGTG CACGACGGCC GTGAGCGGGT
 29521 GCGCCGGGTC GATCCCCGCC AGTACGGAGG CGAGTTCGTC CGGGTGGCG ACGTCCGAGG
 60 29581 CGATCGCCGT GACCTCGGCG CGGGCACGT CGTCGCCGT CGCGCTGCCG GACAGCATCA
 29641 CGAGCCGGCG CACGCCGTGG CGTTGACGA GGTGGCGGT GATGATGCCG GCCAGCGTCC
 29701 CGGAGCCACC GGTGACGAGC ACGGTGCCGT CGGGTGCAG GCGCGGAGCG TCACCCGCCG
 29761 GGACCGCCGG GGCGAGACGG CGGGCGTACA CCTGGCCGTC ACGCAGCACC ACCTGGGGCT
 29821 CATCGAGCGC GGTGGCCGCT CGGAGCAGCG GCTCGGGCGT GTCCGGGGCG CGGTGACGCA

29881	GGACGATCCG	GCCGGGGTGT	TCGGCCTGCG	CGGTCCGCAC	CAGTCCGGCG	GCCGCGGCCG	
29941	ACCGCAGACC	GGGCCCCGTG	TGGACGGCCA	GGACCGCGTC	GGCGTACCGG	TCGTCGGTGA	
30001	GGAAGCGCTG	CACGGCGGT	AGGACGCCGG	CGCCCAGTTC	CGGGGTGTCG	TCGAGCGGGG	
30061	CACCCCCGCC	GCCGTGCGCG	GGGAGGATCA	CCACGTCCGG	GACCGTCGGG	TCGTCGAGGC	
5	30121	GGCCGGTCGT	CGCGTCTG	GGCGGCAGCT	CGGGGAGCTC	GGCCAGCAC	GGCGCAGCA
	30181	GGCCCGGAAC	GGCTCCCGTG	ATCGTCAGGG	GGCGCCTGCG	CACGGCGCCG	ATGGTGGCGA
	30241	CGGGCCCGCC	GGTCTCGTCC	GCGAGGTGTA	CGCCGTCA	GGTACGGCGG	ACCGCTACCG
	30301	CCGTGGCGCC	GGTGGCGTGG	ACCGGGACGT	CGTCGAACGC	GTACGGAAGG	TGGTCCCCCTT
10	30361	CCGCGGCGAG	GCAGGAGTGC	GGCCGAGCA	GCGCCGGTG	CAGGCCGTAC	CGTCCGGCGT
	30421	CGGCGAGCTG	TCCGTCCGCG	AGGGCCACTT	CCGCCAGAC	GGCGTCGTG	TCGGCCCAGA
	30481	CGGCGCGCGG	GCAGGGCCAGC	GCAGGGCCCGT	CCGTGTACCC	GGCTCGGGCC	AGACGGTCGG
	30541	CGATGTGTC	GGGGTCCACC	GGCCGGGCGG	TGGCGGGCG	CCACGTGAC	GGCATCTCCC
15	30601	GCACGGCCGG	GGCCGTCCGC	GGGTGCGGGGG	CGAGGATTCC	GTGCGCGTGC	TCGGTCCACT
	30661	CCCCCGCCGC	GTGCCGCGTG	TGACCGGTGA	CCGCGCGCG	GCCGTCCGCC	CCGGCGCCCG
	30721	TCACCGTGAC	GGAGAGCGCG	AGCGCACCGG	ACCGCGGCAG	CGTGAGGGGG	GTGTCACCGG
	30781	TGAACGTGTC	GAGGGCGCCG	CAGCCGGCTT	CGTCGCCCCG	CCGGATCGCC	AGATCCAGGA
	30841	GGCCCGCGGC	GGCAGCACC	GCGAGGCCGT	GCAGGGAGTG	CGCCAGCGGA	TCGGCGCGT
20	30901	CGACCCGGCC	GGTGAGCAC	AGGTCGCGCG	TGCGGGCAG	GGTGACCGCC	GCGGTCAAGCG
	30961	CCGGGTGCGC	GACCGCGTC	TGTCGGGCCG	GGGCGCGTC	GCCCGCGGT	TGGGTGCCGA
	31021	GCCAGTAGCG	GACCCGCTCG	AACGGGTACCG	TGCGGGGTG	CGAGGCGCGT	GCCGGCGCGG
	31081	GGTCGATGAC	CTTCGGGCC	TGACCGGTGA	CGCCGTGCGT	GTGCGGCCGG	GCGAGCGCGG
	31141	TCAGGGCGGA	TCGGGGTTC	TGCGTCCGGT	GCAGCATCGG	GATGCCGTG	ACGAGTCGGG
	31201	TCAGGCTCCG	GTCCGGGCCG	ATCTCCAGGA	GCACGGCCCC	GTGCGGCCG	GCGACCTGTT
25	31261	CCCCGAACCG	GACGGTGTG	CGGACCTGTC	GTACCCAGTA	CTCCGGCGT	GTGCAGGCCGG
	31321	CGCCCGCGGC	CATCGGATC	CTCGGCTCGT	GGTACCGTAC	GCTCTCCGCC	ACCTTCGCGA
	31381	ACTCCTCGAG	CATCGGCTCC	ATCCGGCCCG	AGTGGAACCG	GTGGCTGGTC	CCGAGGCCGG
	31441	TGAAGCGGCC	GAGCCGGCC	GCGACGTCGA	GCACCGCCTC	CTCGTACCCG	GAGACACGA
	31501	TCGACGCGGG	CCC GTT GACC	GCAGCGATCT	CCACGCCGTC	CCG CAG CAG C	GCGAGCGCGT
30	31561	CCCGTTCCGA	CGCGATCACG	GGGGCCATCG	CCCCGCCGGA	CGG CAG CGCC	TGCATCAGGC
	31621	GGGCCCGTGC	GGACACCAGC	CTGACCGCGT	CCTCCAGGG	CCAGACGCCG	GCGACGTACG
	31681	CGGGGCCAG	CTCGCCGATC	GAATGGCCCA	CGAAGGCGTC	CGGGCGTACG	CCCCACGCC
	31741	CGAGCTGTGC	GGCGAGTGCG	ACCTGGAGCG	CGAACACCGC	GGGCTGGCG	TACCCGGTGT
	31801	CGTGGAGGTC	GAGCCGGCG	GGCACGTCGA	GGGCGTCCAG	CACCTCGCCG	CGAGTGC
35	31861	CGAAGACGTC	GTAGGGCGCG	GCCAGTCCGT	CGCCCATGCC	GGGACGTTGT	GAGCCC
	31921	CGGAGAAGAG	CCACACGAGG	CGGCGGTCCG	GTTCTCGGGC	GCCGGTGACC	GTGTCGGTGC
	31981	CGATCAGCGC	GGCCC GGTC	GGGAAGGCCG	TGCGGGCGAG	CAGGGCCGCG	GCCACCGCGC
	32041	GCTCGTCCCTC	CTCGCCGGT	GCGAGGTGGG	CGCGCAGGCC	GTGTACCTGT	GCGTCAGTG
	32101	CCTGCGGGGT	GC GTG CCG AG	AGCAGCAGGG	GCAGCGGTCC	G GTG TCG GGT	GCCGGGGCGG
40	32161	GTTGGGGG	CGGTGGGGG	TGGCTTCGA	GGATGATGTG	AGC GTT GGT	CCGCTAACGC
	32221	CGAAGGAGGA	CACCCGGCG	CGCCGTGGGC	GGTCGGTTTC	GGGCCAGGGG	CGGGCGTCGG
	32281	TGAGGAGTTC	GACGGCGCCG	GCCGTCCAGT	CGACGTGCGA	GGACGGCGT	TCCACGTGCA
	32341	GGGTGCGCGG	CAGGGTGC	TGCGCATGG	CGAGGACCAT	CTTGATGACA	CCGGCGACGC
	32401	CGCGGGCGGC	CTGAGTGTG	CCGATGTTGG	ACTTCAGCGA	GCCCAGCAGC	ACCGGGGTGT
45	32461	CGCGATGCTG	CCCGTAGGTG	GCCAGTACCG	CCTGCCCTC	GATGGGGTCG	CCCAGCCTGG
	32521	TCCCGGTGCC	ATGCGCTCG	ACAGCGTCA	CATCCGCCG	GGTGAGGCCG	GCGTTGGCCA
	32581	GCGCCTGCCG	GATCACCCGC	TCTCGCAGC	GCCCGTCCG	CCCGGACAAC	CCGTTGGAAAG
	32641	CACCGTCCG	GTTGACCGCC	GAACACCGCA	CGACCGCCAG	GACATTGTGG	CCGTGCCGCT
	32701	CGGGCTCGGA	GAGCCTCTG	ACGATCAGCA	CACCGGATCC	CTCGCGA	CCGGTGCACAT
50	32761	CAGCCGCATC	CGCGAACGCC	TTGCAGCGC	CGTCCCCGG	GAGGCCCCCG	TGCTGGGAGA
	32821	AGTCCACGAA	GCCGGACGGC	GAGGCCATCA	CGT GAC GCG	GCCGACCACG	GCGAGCGAGC
	32881	ACTCCCCCGA	GGCGAGGCC	TGCCCGGCC	GGTGCAGCGC	CACCAAGCGAC	GACGAACACG
	32941	CCGTGTCCAC	CGTGACCGCC	GGACCCCTCA	AACCGTAGAA	GTACGACACG	CGACCGGACA
	33001	GCACACTGGT	CTGGGTGCTG	GTGGCACCGA	AACCGGCCG	GT CGG CTCCA	GTGCCG
55	33061	CGTAGAAGTA	GCCGCCCATG	AACACGCCGG	TGTCGCTTC	GCGCAGCGAC	TCCGGGAGGA
	33121	TCCCGGCGT	TTCCAGCGCC	TCCCACGAGG	TCTCCAGGAC	CAGACGCTG	TGCGGGTCCA
	33181	TCGCCAGCGC	CTCACCGGA	CTGATCCC	AGAACGCCG	GTCGAAGTCC	GCCACCCCGG
	33241	CGAGGAAGCC	ACCATGACG	ACGGTCGAGC	TGCGGGATG	ATCCGGATCG	GGATCGTACA
	33301	GCCCCGCCAC	GTCCCACCA	CGGTCCGTG	GAAACGCCGT	GATCCCGTCA	CCACCCGACT
60	33361	CCAGCAGCG	CCACAAGTCC	TCCGGCGAGC	CGACCCACC	CCG CAG CGCG	CAGGCCATCC
	33421	CCACGATCGC	CAACGGCTCG	TCCGTCCGGA	CGGCCGCGGT	CGGGGTACCG	CGCCGGGTGG
	33481	TGGCCCGCGC	GCGGCCAGT	TCGTCCAGG	GGGCGCGCAG	CGCCTGCGCC	GTGGGGTGGT
	33541	CGAAGACGAG	CGTAGCGGGC	AGCGTCAGGC	CCGTCGCGTC	GGCCAGCGCG	TTGCGCAGTT
	33601	CGACGCCGGT	CAGCGAGTCG	AAGCCCAC	CCCTGAACGC	GCGCGCGGGT	GCGATGGCGT
	33661	GGCGTGC	GTGGCCGAGC	ACCGCGGCAG	CGCTGGTACG	GACGAGGTG	AGCATGTCG

33721	GCGCGGCCGG	AGGTGCGGAC	GTGCGCCGGA	CGGCCGGCAC	GAGGGTGCCT	AGGACCGGCG	
33781	GGACCCGGTC	GGACCGGGCG	ACGGCGGCGA	GGTCGAGCCG	GATCGGCACG	AGCGCGGGCC	
33841	GGTCGGTGTG	CAGGGCCGCG	TCGAACAGGG	CGAGCCCCTG	TGCGGCCGTC	ATCGGGGTCA	
33901	TGCCGTTGCG	GGCGATGCGG	GCCAGGTCGG	TGGCGGTCA	CCGCCCCGCC	ATCCCGTCCG	
33961	CCGCCTCCCA	CAGTCCCCAG	GCGAGCGAGA	CGGGGGCGAG	CCCCTGGTGG	TGCCGGTGGC	
34021	GGCGAGCGC	GTCGAGGAAC	GCCTTGCCTGG	TCCGCTAGTT	GGCCTGACCC	GCGCCGCCGA	
34081	ACGTGGCGA	TATGGACAG	TACAGGACGA	ACCGGCCAG	GTCGAGATCG	CGCGTCAGCT	
34141	CGTGCAGGTG	CCAGGCGACG	TCCGCTTGA	CCCGCAGCAC	GCGTCCCAC	TGCTCCGGC	
34201	GCATGGTCGT	CACGGCCGCG	TCGTCGACGA	TCCCAGGCAT	GTGACGACG	GCGCGCAGCC	
10	34261	GCTGGCGAC	GTCGGCGACG	ACTGCGGCCA	GCTCGTCGCG	GTCGACGACG	TCGGCGGCCA
34321	CGTACCGCAC	GCGGTCTCC	TCCGGCGTGT	CGCCGGGCCG	GCCGTTGCGG	GACACCAACGA	
34381	CGACCTCGGC	GGCCTCGTGC	ACGGTGAGCA	GGTGGTCCAC	GAGGAGGCGG	CCGAGCCCGC	
34441	CGGTGCCGCC	GGTGACGAGG	ACGGTCCCAG	CGGTCAGCGG	GGAGGTTCCG	GTGGCCGCCG	
34501	CGACACGGCG	CAGACGGGCC	GCACGCGCTG	TGCCGTGGC	GACCCGGACG	TGCGGCTCGT	
15	34561	CGCCGGCGGC	GAGCCGGCC	GCTATGGCGG	CGGGCGTGT	CTCGTCCGCT	TCGATCAGCG
34621	CGACCGGGCC	GGGATGCTCC	GTCTCCGCCG	TCCGGACAG	GCCGCCGAGC	GCTTCCTGCG	
34681	CGGGATCGCC	GGTACGGGTG	GCCACGATGA	GCCGGGATCG	CGCCCAGCGC	GGCTCGGCCA	
34741	GCCAGGTCTG	CACGGTGGTG	AGCAGGTGCG	GGCCCAGCTC	CCGGGTCCCG	GCGCCGGGCG	
34801	AGGTGCCCGG	GTCGCGGGT	TCCACGGCCA	GGACACGAC	CGGGGGGTGC	TCGCCGTGG	
20	34861	GCACGTCGGC	GAGGTACGTC	CAAGTCGGGA	CGGGTGACGC	GGGCACGGGC	ACCCAGGCGA
34921	TCTCGAACAG	CGCCTCGGCA	TCGGGGTCGG	CGGCCCCAC	GGTCAGGCTG	TCGACGTCAA	
34981	GGACCGGTGA	GCCGTGCTCG	TCCGTGGC	CGATGCGGAC	CATGTCGGGG	CCGACCGCGT	
35041	CCAGCAGCAC	GCGCAGCGC	GTCGCGGCCG	GCGCGTGGAT	CCTCACGCCG	GACCAAGGAGA	
35101	ACGCCAGCCG	GCGCCGCTCC	GGGTCCGTGA	AGACCGTCCC	GAGGGCGTGC	AGGGCCCGT	
25	35161	CGAGCAGCAC	GGGGTGCAGC	CCGTACCGGG	CGTCGGTGT	CTGTTGGC	AGGCGGACCG
35221	ACCGTAGGCG	GCGGCCCTCC	CCCGTCCACA	TCGCGGTAT	GGCCCGGAAAC	GCGGGCCCGT	
35281	ACGAGAGCGG	CAGCGCGTCG	TAAGAGCCGG	TCAGGTGCGC	CGGGTCCGGC	TCGGCGGGCG	
35341	GCCAGTCCAC	GGGCTCCGCC	GGACCGCCAG	TGTCCACGCT	CAGCGCTCCG	GTCGCACTGAA	
35401	GCGCCCAGGG	GCCCCTGCCG	GTACGGCTGT	GCAGACTCAC	CGACCCGCGT	CCGGACACCT	
30	35461	CGGTTCCGAC	GGTGGCTCTGG	ATTCCTGTG	CGCCGTGCGC	GTCGACCAACC	ACCGGGCGCA
35521	CGATGGTCAG	CTCCCGATC	TCCCGCTGTC	CGAGCGGGGC	TCCCGCTTCG	GCGAGCAGTT	
35581	CCACGAGCGC	CGAGCCGGGC	ACGATGACCC	GGCCGTCAC	CTCGTGGTCG	GCGAGCCAGG	
35641	GCTGACGGCG	TACCGAGACA	CCCGGGTGGC	CAGCGCCGCC	TCGCCGTGCG	GCGAGGTCGA	
35701	CCCACGAGCC	GAGCAGCGGG	TGGCCGGACG	TTCCCGCCGG	TTCCCGTGTG	ATCCAGTAC	
35	35761	GGTCACGGCG	GAACGGGTAC	GTGGCAGCG	GCACCAACCG	ACCGTGTGCG	AACGACJAGG
35821	TGACGGGCAC	GCCCCGGACC	CAGAGCGCGG	CGAGCGACCG	AGTGAAGCGG	TCCAGGCCG	
35881	CCTCGCTCG	CCCGAGTGTG	CCGGTGACGA	CCGTATGCGC	ATGCCCGGCG	AGCGTGTCT	
35941	CCAGTGCAGG	GGTGAGCACG	GGATGCGCGC	TGACCTCGAC	GAACGCGCGG	TATCCCGGTT	
36001	CCGCCAGGTG	CCCGGTCGCG	CGGGCGAAC	GAACGGTGTG	GCGCAGGTTG	TCGTACCAAGT	
40	36061	AGGCGCGTC	CGCGGGCGG	TCCAGCCACG	CCTCGTCCAC	GGTGGAGAAAG	AACGGGACGT
36121	CGGCGTGC	CGGAGTGTAT	CCGGCGAGAG	CGTCGAGCAG	CGCGCCGCGG	ATCGTTTCGA	
36181	CATGCGCGGT	GTGCGACGCG	TAGTCGACGG	CGATCCGGCG	GGCGCGGGGG	GTGGCGGCCA	
36241	GCAGCTCCTC	CACGGCGTCG	GCCGCACCGG	CGACAAACGAT	CGACGGGGGT	CCGTTGACCG	
45	36301	CGGGCACCTC	CAGGGCCCCG	GCCCACACGG	CGGCGTCGAA	GTCGGCGGGC	GGCACCGAGA
36361	CCATGCCGCC	CTGCCCCGCC	AGTTGGTGG	CGACGAGTCG	GCTGCGCACC	GCGACGACCT	
36421	TCGCGCGTC	GTCCAGGGTG	AGCACCCCGG	CGACGCAAGGC	CGCGCGCAGT	TCGCCCTGGG	
36481	AGTGGCCGAC	GACCGGGGCC	GGGGCGACCC	CGTGCAGCAG	CCACAGCTCC	GCCAGCGCCA	
36541	CCATCACCGC	GAACGACGCG	GGCTGCACGA	CATCGACCCG	GTCGAACGCG	GGCGCTCCGG	
50	36601	GCCGCTGGGC	GATGACGTCC	AGCAGGTCCC	ATCCGGTGTG	CGGGCGAGC	GCCGTGGCGC
36661	ACTCGCGGAG	CCGCCGGGCG	AAACACGGCT	CGGTGGCGAG	CAGTTGGCA	CCCATGCCGG	
36721	CCCACTGGGA	GCCCTGCCG	GGGAACGCGA	ACACGACACG	TGTGTCGGT	ACGTCGGCCG	
36781	TTCCCGTCAC	GGCCCCCGGC	ACTTCGGCAC	CACGGCGAA	CGCCTCCGCC	TCTCGGGCCG	
36841	GCACGACCGC	CCGGTGGCGC	ATGGCCGTCC	GGGTGGTGC	GAGCGAGTGG	CCGACCGCG	
36901	CCGGCGCGCC	AGTGAGCGGG	GCCAGCTGTC	CCGCGACGTC	CCGCAGTCCC	TCCGGGTCTC	
55	36961	GGGCCGACAT	CGGCCAGAC	ACGTCTCGG	GCACCCGCTC	GGCTTCGGGT	GCGGACACGG
37021	GTGCGGGCGC	GGCGGGGGGC	CCGGCCTCCA	GGACGACATG	GGCGTTGGTG	CCGCTGATGC	
37081	CGAACGACGA	GACACCCGCA	CGCCGGGCGC	GCCCGGTGAC	CGGCCACGCG	TCACTGCGGT	
37141	GCAGCAGCGC	GATGTCGCCG	TCCAGTCGA	CGTGCCTGGGA	CCGCTCGTCG	ACGTGAGCG	
37201	TGCGCGGCAG	GACGCCGTGC	CGCATCGCCA	TGACCATCTT	GATGACGCCG	GCGACGCCGG	
60	37261	CCGGGGCGCTG	GGTGTGGCCG	ATGTTCGACT	TGAGCGAGCC	GATCAGCAGC	GGATGACACGC
37321	GTTCGCGCCC	GTAGGCCACT	TGCAAGGCCT	GGGCCTCGAC	GGGGTGCAGCG	AGACGGGTG	
37381	CGGTGCCGTG	TGCCCTCACG	GCGTCGACGT	CACCCGGCGC	CAGGCCGGCG	TCGGCGAGCG	
37441	CACGCTGGAT	GACGCGCTGC	TGCGCAGGCC	CGTTGGGGC	GGACAGCCCC	TTCGACGCGC	
37501	CGTCGGAGTT	GACCGGGAG	CGCGGCACCA	GCGCCAGCAC	GGGGTGGCCG	TGGCGGGTGG	

37561 CGTCGGAGAG CCGCTCCAGC ACCAGGACAC CGGCGCCCTC GGCGAAGCTC GTGCCGTCCG
 37621 CGGTGTCCGC GAAGGCCTTG GCACGGCGT CGGGGGCGAG CCCGCGCTGC CGGGAGAACT
 37681 CGACGAACCC GGTCGTCGTC GCCATCACCG TGACACCGCC GACCAGGGCG AGCGAGCACT
 37741 CCCCCGAGCG CAGCGACCGC GCGGCCTGGT GCAGCGCCAC CAGCGACGAC GAACACGCCG
 5 37801 TGTGACGGT GACCGACGGG CCCTCCAGAC CGAAGTAGTA CGAGAGCCGC CCGGAGAGAA
 37861 CGCTGGTCGG CGTGCAGGTC GCCCCGAAAC CGCCAGGTC CACGCCCGCG CCGTAGCCCT
 37921 GGGTGAACGC GCCCCATGAAT ACGCCGGTGT CGCTGCGCG GACGCTTCG GGCAGGATGC
 37981 CCGCTCGTTC GAACGCCCTC CACGACGCTT CGAGGACCAG ACGCTGCTGC GGGTCCATCG
 38041 CCAGCGCCCTC ACGCGGGCTG ATCCCAGA ACGCGGCCTC GAAGTCCGCG GCGCCGGTGA
 10 38101 GGAAGCCGCC GTGACGCACG GAAACCTTGC CGACCGCGTC GGGGTTGGG TCGTAGAGCG
 38161 CGGCGAGGTC CCAGCCGCGG TCGGGGGGA ACTCGGTGAT CGCGTCCCCG CCGGAGTCGA
 38221 CCAGCCGCCA CAGGTCCTCC GGTGACCGCA CGCCACCGGG CATCCGGCAC GCCATGGCCA
 38281 CGATCGCCAG CGGCTCGTTC CCCGCCACCG TCGGTGCGGG CACTGTGCC GCGGGAGCGG
 15 38341 CAGGGGCCGG CTCACCCCGC CGTTCCCAT CCAGGCCGGC GGCAGGCCG GCGGGTGTGCG
 38401 GGTGGTCGAA GACGGCGTC GCGGAGAGGC GTACCCCGT CGTCTCGCG AGGCTGTTGC
 38461 GCAACCAGGAC ACCGCTGAGC GAGTCGATGC CGAGGTCTT GAACGCCCTC GTGGCGTGA
 38521 TCTCGGAGGC GTCGGCGTGG CCGAGCACGG CGGCCGTGGC CGCACACACG ATGCCAGCA
 38581 GGTACGATC GCGGTCGCGG TCGCGGTGCG GTTGTCTC CGCACGGCG GCGATGCGCG
 38641 GTCGGTCCG CTGCGGAGC GGCTCGGTGG GAATCGCCGC GACCATGAAC GGCACGTCCG
 20 38701 CGCGCAGGCT CGCGTCGATG AAGTGGGTG CCTCGGCCCTC GGTGAGCGGC CGGAACCCGT
 38761 CGCGCACCCG CTGCGGTGCG GCGTCGTCAA GTTGTCCGGT GAGGGTGTGCG GTGGTGTGCC
 38821 ACATGCCCA GGCATGGAG GTGGCGGGTT GGCGAGGGT GTGGCGGTGG GTGGCGAGGG
 38881 CGTCGAGGAA GGCCTGGCG GCGCGTAGT TTCTTGTCC GGGGCTGCCG AGGACGGCGG
 38941 CGGCCTGGA GTAGAGGACG AAGTGGGTGA GGGGTTGGG TTGGGTGAGG TGGTGCAGGT
 25 39001 GCCAGGCCG GTTGGTTTG GGGTGGAGGA CGGTGGTGAG GCGGTGGGG GTGAGGGCGT
 39061 CGAGGATGCC GTCGTCGAGG GTGGCGGGG TGTGGAAAGAC GCGGGTGAAGG GTTGGGGGAA
 39121 TGTGGCGAG GGTGGTGGCG AGTTGGTGGG GGTGCCGAC GTCGAGGGG AGGTGGGTGC
 39181 CGGGGGTGGT GTCGGGGGGT GGGTGCAGG AGAGGAGGTA GGTGTGGGG TGGTTCAGGT
 39241 GGCAGGCCAG GATGCCGGCG AGGGTGGGG AGCCGCCGGT GATGATGATG GCGTGTCCG
 30 39301 GTTGGAGGGG GGTGGTGGTG GGTGGGGTGG TGGTGTGGAG GGGGGTGAAGG TGGGTCGGT
 39361 GGAGGGTGTG GTGGGTGAGG CGGAGGTGGG GGTGGTCGAG GTGGCGAGT TGGGCCAGGG
 39421 GGAGGGGAGT GTGGGGGTGG TCGGTTTCCA TGAGGGCGAT GCGGTGGGG TGTTCTTCT
 39481 GGGCGGTGCG GGTGAGGCCG GTGACGGTGG CGCCGGCGGG GTCGGTGGTG GTGTGGACGA
 39541 TGAGGGTGTG GTCGGTGGTG GTGAGGTGGT GTTGCAGGGC GGTCAAGGAC CGGGTGGCG
 35 39601 GGGTGTGGG GCGGGTGGGT ATGCTCTCGG GTGCGTCGGG GTGGCGGGG GTGATCAGGA
 39661 CGTGTCCCTC GGGCAGGTCA CCGTCGTAGA CGCCTCGGC GACCGCGAGC CACTCCAACC
 39721 GGAGCAGGGT CGGCCCCGAC GGGGTGTCCC CCCGCTCCCT CAGCACCAAGC GAGTCCACCG
 39781 ACACGACAGG ACGGCCATCC GGGTCGGCCA CGCGCACGGC GACGCCGGCC TCCCCCGGG
 39841 TGAGGGCGAC GCGCACCGCG GCGGCCCCCG TGGCGTTCA GCGCACGCC GTCCAGGAGA
 40 39901 ACGGCAGCTC GATCCCCCGC CCCGCGTCGA GGCGCCCGGC GTGCAGGGCC GCGTCGAGCA
 39961 GTGCCGGATG CACACCGAA CGTCCCGCTC CGCGGGCTG CTCGTCGGGC AGGCCACCT
 40021 CGGCATACAC GGTGTCACCA TCACGCCAGG CAGCCCGCAA CCCCTGGAAC GCCGACCCGT
 40081 ACTCATAACC GGCATCCCGC AGTTCGTCAT AGAACCCGA GACGTCGACG GCGCGGCCG
 40141 TGGCCGGCGG CCACTCGCAG AACGGCTCAC CGGAAGCGTT GGAGGTATCC GGGGTGTCCG
 45 40201 CGGTCAAGGT GCCGCTGGCG TGCCGGGTCC AGCTGCCGT GGCCTCGGTA CGCGCGTGG
 40261 CGGTCAACCGG CGGCCGTCCG GCCTCATCGG CCCCTCCAC GGTCAACCGAC ACATCCACCG
 40321 CTGCGGTAC CCGCACCCAG AGCGGGGATT CGATGACCAAG TTCATCCACC ACCCCGCAAC
 40381 CGGTCTCGTC ACCGGCCCGG ATGACCAAGCT CCACAAACGC CGTACCCGGC AGCAGAACCG
 40441 TGCCCCGAC CGCGTGTACA GCCAGCCAGG GATGCGTACG CAATGAGATC CGGCCGGTGA
 50 40501 GAACAACACC ACCACCGTC TCGGGGGCA GTGCTGTGAC GGCAGGCCAGC ATCGGATGCG
 40561 CGGCCCGGT CAGCCCCGGC GCGGACAGGT CGGTGGCACC GGCGCCCTCC AGCCAGTACC
 40621 GCCTGTGTC GAACGCGTAG GTGGGCAGAT CCAGCAGCCG CCCCCGGCACC GTTCGACCA
 40681 CGGTGCCCCA GTCCACCCCCC GCACCCAGAG TCCACGCCCT CGCCAACGCC CCCAGGCCACC
 40741 GCTCCCAGCC ACCGTACCA GTCCCGAACG AGCCACCGT GCGGGCCTGT TCCATCGCCG
 55 40801 GCAGCAGCAC CGGATGGGA CTGCACTCCA CGAACACCGA CCCGTCCAGC TCCGCCACCG
 40861 CGGCATCCAG CGCGACAGGG CGACCGAGGT TCCGGTACCA GTACCCCTCA TCCACCGGCT
 40921 CGGTACCCCA GGCCTGTTC ACGGTCGAC ACCACGCCAC CGACCCGGTC CGGCCGGAAA
 40981 TTCCCTTCAG TACCTCAGCG AGTTCGTCCT CGATGGCCTC CACGTGAGGC GTGTGGGAGG
 41041 CGTAGTCGAC CGCGATACGA CGCACCCCGA CCCCATCAGC CTCATACCCG GCCACCACT
 60 41101 CCTCCACCGC CGACGGGTCC CCCGCCACCA CGTCAAGGC CGGACCATTA CGCGCCGCGA
 41161 TCCACACACC CTCGACCAAGA CCCACCTCAC CGGCGGGCAA CGCCACCGAA GCCATCGCCC
 41221 CCCGGCCGGC CAGCCCGGCC GCGATCACCC GACTGCGCAA CGCCACCAAG CGGGCGGGCG
 41281 CCTCCAGGCT GAGGGCTCCG GCCACACACG CGGCCCGGAT CTCCCCCTGC GAGTGTCCGA
 41341 CCACAGCGTC CGGCACCGAC CCATGCGCCT GCCACAGCGC GGCCAGGCTC ACCCGCACCG

41401 CCCAGCTGGC CGGCTGGACC ACCTCCACCC GCTCCGCCAC ATCCGACCGC GACAACATCT
 41461 CCCGCACATC CCAGCCCGTG TCGGGCAACA ACGCCCGCGC ACACTCCTCC ATACCGAGCCG
 41521 CGAACACCGC GGAACGGTCC ATGAGTTCCA CGCCCATGCC CACCCACTGG GCACCTGCG
 41581 CGGGGAAGAC GAACACCGTA CGGGCTGTAT CCACCGCCAC ACCCATCACCG CGGGCATCAC
 5 41641 CCAGCAGCAC CGCACGGTGA CGGAAGACAG CACGCTCACG CACCAACCCC TGCGCGACCG
 41701 CGGCCACATC CACCCCAACCC CGCGCAGAT ACCCCCTCCAG CCGCTCCACC TGCCCCCGCA
 41761 GACTCACCTC ACCACGAGCC GACACCGGCA ACGGCACCAA CCCATCACCA CCCGACTCCA
 41821 CACGCGACGG CCCAGGAACA CCTCTCAGGA TCACGTGCGC GTTGTACCG CTCACCCCCGA
 41881 ACGACGACAC ACCCGCATGC GTTGCCCGAT CCGACTCGGG CCACGGCCCTC GCCTCGGTGA
 10 41941 GCAGCTCCAC CGCACCGGCC GACCAGTCCA CATGCGACGA CGGCTCGTCC ACGTGCAGCG
 42001 TCTTCGGCGC GATCCCATGC CGCATCGCCA TGACCATCTT GATGACACCG GCGACACCCG
 42061 CAGCCGCGCTG CGCATGACCG ATGTTGACT TGACCGAAC GAGGTAGAGC GGCCTGTCGC
 42121 GGTCTGCCC GTAGGCCGCG AGGACGGCCT GCGCCTCGAT CGGGTCGCCC AGCCCGTGC
 42181 CGGTGCCGTG CGCCTCCACC ACgtCCACAT CGGCGCGCG CAGTCCGGCG TTGACCAACG
 15 42241 CCTGCCGGAT CACGCGCTGC TGGCGACGC CGTTGGGGC GGACAGTCCG TTGGAGGCAC
 42301 CGTCCTGGTT CACCGCCGAG CGCGGAGCGA CGCGAGAAC GGTGTGCCG TTGCGCTCGG
 42361 CGTCGGAGAG CCGCTCCAGC ACGAGAACGC CGACGCCCTC GGCGAAGCCG GTCCCGTCCG
 42421 CGCGTCCGGC GAACGCCCTG CACCCTCGT CGGGGAGAG TCCGCGCTGC CGGGAGAACT
 42481 CCACGAGCTC TGCGGTGTT GCCATGACGG TGACACCGCC GACCAGCGCC AGGGAGCACT
 20 42541 CCCCGGCCCG CAGTGCCTGT GCCGCCTGGT GCAGGGCGAC CAGCGACGAC GAGCACGCCG
 42601 TGTCGACCGT GACCGCCGGG CCTGAAAGTC CGTACACGTA CGAGAGGCGC CGGGACAGGA
 42661 CGCTCGTCTG CGTCGCCGTG ACACCGAGCC CGCCCAAGGTC CGGGCCGACG CGTAGCCCT
 42721 GGTGAAACGC GCCCATGAAC ACGCCGGTGT CGCTCTCCCG GAGCCTGTC GGCACGATGC
 42781 CGCGCTTCTC GAACGCTCC CAGGAGGTCT CCAGGATCAG GCGCTGCTGG GGGTCCATCG
 25 42841 CCAGCGCCTC GTTCGGACTG ATGCCGAAGA ACACGGCGTC GAACCCGGCG CGGGCCAGGA
 42901 ATCCGCCGTG GCGTGTGCG GAGCGGCCGG CGCGTCCCG GTCCGGTCTG TACAGCGCTG
 42961 CGACGTCCA GCGCGGTG GTGGGAACT CGGTGATCG CTCGGTACCG CGGGCGACGA
 43021 GCGGCCACAG GTCCCTCGGC GAGGCGACCC CGCCGGCAG CGCGCACGCC ATGCCGACGA
 43081 TCGCGACGGG GTCGCCGGAG CGGAGGGTCT GGGCGTCTG GGGTCCGCT GTCGCGGAGC
 30 43141 CGGGGAGGTG GGCACGGAAC GCACGGCGAG TGGGGTGGT GAACCGGGTT GACCGGGCA
 43201 CCCCGAGACC CGTCCCGCGC GCGACGGTGT TGGTAACTC GACGGTGGT AGCGAGTCGA
 43261 GGCGTCTC GCGAACGTG CGGTCCGGGG AGCAGTGTCC GGCACCCGGC AGGCCCAGGA
 43321 CGGTGGCGAC GCTGTCCGG ACCAGGTCTGA CGAGTACGTC CTCCCGGCC GCACGGGCCG
 43381 CGGGGAGGCG GTTCGCCAC TCTGTTCGG TGGCGTCCGG CTCGGCCGGT CGGTCACTG
 35 43441 CGGTGAGGAT CGGCGGGTGC CGGCCCGCCA TCGTCGCCGC CGCGCCCGCG CGGAAACGG
 43501 TCCGGGCCAC GATGTACGAG CGGCCGCCCG CGATGCCCT CTCGATCAGG TCGCCGGTCA
 43561 GCGCCGGCCG TTCGATGCC GGCAGCGCGC GGACGGTGC GGTGGGGAGT CCCTCCGCGG
 43621 CCCGTGGCCG GGTGTGGCG TCGCGCCGG CGGGGCCGTC GAGCAGGACG TGCACGAGCG
 43681 CGCCGGGGTT CGCGCTTCC TCGGCTGCCG TGGTCACTG GGTGAGGCGC GTCTCGTGC
 40 43741 GGAGCAGGCC GGCACGGTG TCGCGTCCT CCCCGGTGAC CAGGACCGGC CGTCCCGGGC
 43801 CGATCGGAGG CGGCACGGTG AGGACCATCT TGCCGGTGTG CGGGCGTGG CTACATCCACG
 43861 CGAACCGCTC CGCGCACGG CGGATGTCCC ACGGCTGCAC CGGCAGCGGG CACAGCTCAC
 43921 CGGGTCTGAA CAGGTCTGAGG AGCAGTTCTGA GGATCTCCCG CAGGCGCGCG GGATCCACGT
 43981 CGGCCAGGTC GAACGGTGC TGGCGCGT GGCGGATGTC GGTCTGCCC ATCTGACGA
 45 44041 ACCGGCCGCC CGGTGCGAGC AGGCCGATGG ACACGCTGAG GAGTTCACCG GTGAGCGAGT
 44101 TGAGCACGAC GTCGACCGC GGGAAAGGTGT CGCGAACGC GGCGCTGCG GAGTCGCCA
 44161 CATGGTCTGGT GTCGAAGCCG TCGCGTGCAGA CGAGGTGTT TTTGGCGGG CTGGCGTGG
 44221 CGTACACCTC CGCGCCGAGG TGGCGGGCGA TCCGGTGC CGCCATGCCG ACACGCCCG
 44281 TCGCGCGTG GACCAGGACC TTCTGGCCGG GTCGAGCTC GCCCGCGTCG ACGAGGCCGT
 50 44341 ACCAGGCGGT GGCGAACACG ATGGGCACGG ACACGCGCAGT GGGGAACGAC CATCCCCGTG
 44401 GGATCCGTGC GACCAGGCC CGGTCCCGA CCACGCTGCG CGGGAACCGC TCCTGCACGA
 44461 GACCGAACAC CGGGTCCGGC GGGGCCAGGT CGTCGACGCC GGGTCCGACT TCGGTACGA
 44521 TGCCCGCGGC CTCCCCGCC ATCTGCCC CGCCCGGGTA GGTGCCGAGC GCGATCAGCA
 44581 CGTCGCGAA GTTCAGCCCC GCGGCCGCGA CGTCGATGCG GACCTCGCC CGGGCAGGG
 55 44641 CGCGGGCGGG ACGTCGAGCG GGGCGACGAC GAGGTCGCGG AGCGTTCCGG AGGCGGGCG
 44701 GCGCAGCGCC CACTGGCGG GTCCGGCAGGG GGGTGGTGC CGGGCGTACCG AGCCGGGGCA
 44761 CGTAGGCCAC GCGGCCCGC AGCGCGATCT GGGGTTGCG GAGCGAGGCC CGGGGGGGGA
 44821 CGACGTCTGC ATCGCCGTCC GTGTCACCCA GCACGAACGA TCCGGGTTCG CGGGCCTGGC
 44881 GGCGCAGCGC CTCGTCCCCAG AGCGGGGCCGT GGTCCCGTC CGGGGATCTCG GCGGGGCCGA
 60 44941 CGCCCACCGC GCGCGGGTG ACACGGTCC GGCAGGGTGA CGGGGTGCG GGCAGGTGCG
 45001 GCGCTCCCA GACCAAGTCG CACAGCGTGG CCTCGCCACT GCGGGTGGCG ACCAGATGGG
 45061 CGGGCAGCCC CGCGAGCCGC CGCGCGCTGGA CCTTGGCCGA CGCGGTGCGG GGGATCGTGG
 45121 TGACGTGCCA GATCTGCTG GGCACCTTGA AGTAGGCGAG CGGGCGGGCG CACTCGCGA
 45181 GGATCGCCTC GGCAGGGACG CGGGGGCCGT CGGAAACGAC GTAGAGCACG GGTATGTCG

45241	CGAGGACGGG	GTGCGGGCGG	CCC GCCGCGG	CGGCGTCCCG	GACACCGGGC	ACCTCCTGGG	
45301	CGACGGTCTC	GATCTCCCGG	GGGTGGATGT	TCTCCCCGCC	CGGGATGATC	AGCTCCTTGA	
45361	CCCGGCCGGT	GATCGTCACG	TGTCCGGTCT	CGGCCTGACG	TGCGAGGTCC	CCGGTGCGGT	
45421	ACCAGCCGTC	CACGAGCACC	TGGGCGGTCG	CCTCCGGCTG	GGCGTGGTAG	CCGAGCATGA	
5	45481	GGCTCGGCC	GCTCGCCAC	AGCTCGCCCT	CCTCGCCGGG	TGCCACGTG	GCGCCGGACA
	45541	CCGGGTCGAC	GAACCGCAGC	GACAGGCCCG	GCACGGGCAG	CCCGCACGAG	CCGGGAACCC
10	45601	GCGCATCCTC	CAGGGTGTG	GCCTGGTACG	AGCCGGTCGT	CTCGGTGCAG	CCGTACGTGT
	45661	CGAGCAGGGG	CACGCCAAC	GTCGCCTCGA	AATCCCTGGT	GAGCGACGCC	GGCGAGGTGG
15	45721	ATCCGGCGAC	CAGCGCCACG	CGCAGCGCAG	GAGCCCGCGG	CTCGCCGGAC	ACGGCGCCGA
	45781	GGAGGTAGCG	GTACATCGTC	GGCACGCCGA	CGAGCACCGT	GCTGGAGTGT	TCGGCCAGGG
20	45841	CGTCGAGGAC	GTCACGCCG	ACGAAGCCGC	CCAGGATACG	GGCGGACCG	CCGACCGTGA
	45901	GGACGGCGAG	CAGGCCAGGG	TGGTGGCCGA	GGCTGTGAA	CAGCGGGGCG	GGCCAGAGCA
25	45961	GTTCGTCGTC	CTCGGTACG	CGCCAGGACG	GCACGTGCA	GTGCATCGC	GACCACAGGC
	46021	CGCTCGCTG	TGCGGAAACC	ACGCCCTTG	GACGGCCGGT	GGTGCCTGGAG	GTGTAGAGCA
30	46081	TCCAGGCGGG	TTCTCCAGG	CCGAGGTCTG	CGCGGGCGG	GCACGGCGGC	TCGGTCCCAGG
	46141	CGAGGTCTC	GTAGGAGACG	CAGTCCGGTG	CCC GGCGCC	GACGAGCACG	ACGGTGGCGT
35	46201	CGGTGCCGGT	GGGGCGCAC	TGGTCGAGGT	GGGTTCTCG	GGT GACCG	ACGGTGCAGC
	46261	CGGAGTCCCT	CAGGAAGTGG	GGCAGTTCGG	CGTCCGGCGC	GTCCGGTTG	AGCGGGACCG
40	46321	CGACGGCGGC	GGCGCGGGCG	GCGCGAGGT	AGACCTCGAT	GTTCTCGATC	CGGTTGCCGA
	46381	GCAGCATCGC	GACCCGGTCG	CCGCGGTCA	CGCCGGACGC	GGCGAGGTGT	CCGGCGAGCC
45	46441	GGCCGGCCCG	GAGCCGGAGT	TGCGTGTACG	TCACGGCGC	TTGGGAATCC	GTGTAGGCGA
	46501	TCCGGTCGCC	GGCTCGCTG	GCATGGATGC	GGAGCAATT	GTGCAACGGC	CGGATTGGGT
50	46561	CCACACGCGC	CATGGAAACA	CCTTCTCTC	GACCAACCGC	ACAACAGCAC	GGAAACCGGCC
	46621	ACGAGTAGAC	GCCGGCGACG	CTAGCAGCGT	TTTCCGGACC	GCCACCCCT	GAAGATCCCC
55	46681	CTACCGTGGC	CGGCCCTCCCC	GGACGCTCAT	CTAGGGGGTT	GCACGCATAC	CGCGTGCAGT
	46741	AATTGCCTTC	CTGATGACCG	ATGCCGGACG	CCAGGGAAGG	GTGGAGGCGT	TGTCCATATAC
60	46801	TGTCACGGCG	CCGTATTGCC	GCTTCGAGAA	GACCGGATCA	CCGGACCTCG	AGGGTGACGA
	46861	GACGGTGCTC	GGCCTGATCG	AGCACGGCAC	CGGCCACACC	GACGTGTGCG	TGGTGGACGG
	46921	TGCTCCCCGG	ACCGCCGTG	ACACCACGAC	CCGTGACGAC	GAGGGCTTCA	CCGAGGTCTG
	46981	GCACGCACAG	CGCCCTGTG	AGTCCGGCAT	GGACAAACGGC	ATCGCCTGGG	CCCGCACCGA
	47041	CGCGTACCTG	TTCGGTGTCG	TGCGCACCGG	CGAGAGCGGC	AGGTACGCGG	ATGCCACCGC
	47101	GGCCCTCTAC	ACGAACGTCT	TCCAGCTCAC	CCGGTCGCTG	GGGTATCCCC	TGCTCGCCCG
	47161	GACCTGGAAC	TACGTAGCG	GTATCAACAC	GACGAACGGC	GACGGGCTGG	AGGTGTACCG
	47221	GGACATTCTGC	GTGGGCCGCG	CCCAGGCGCT	CGACGAGGGC	GGGATCGACC	CGGCCACCAT
	47281	GCCCGCGGCC	ACCGGTATCG	GGCCCACGG	GGCGGCATC	ACCTCGCTG	TCCTCGCCGC
	47341	CGGGGGCGGA	GTGCGGATCA	ACATCGAGAA	CCCCGCCGTC	CTCACGGCCC	ACCAACTACCC
	47401	GACGACGTAC	GGTCCGGCGC	CCCCGGTCTT	CGCACGGGCC	ACCTGGCTG	GCCCGCCGCA
	47461	GGGGGGCCGG	CTGTTCATCT	CCGCGACGGC	CGGCATCCTC	GGACACCGAA	CGGTGCACCA
	47521	CGGTGATGTG	ACCGGCAGT	GGCAGGTGCG	CCTCGACAA	ATGGCCCGGG	TCATCGCGC
	47581	GGAGAACCTG	CGGCGCACG	GGTCCAGCG	GGGGCACGTC	CTCGCCGACG	TGGACCACCT
	47641	CAAGGTCTAC	GTCCCGCGCC	GGCAGGGATCT	CGATACGGTC	CGCCGGGTCT	GCGCGCACG
	47701	CCTGTGAGC	ACCGCGGCCG	TGCCCTTTT	GCACACCGAC	ATAGCCCGCG	AGGATCTGCT
	47761	CGTCGAAATC	GAAGGCATGG	TGGCGTACGAA	ATACCGGTA	AAAGGCCCCG	GACGCTGCGC
	47821	CTCGGCGGAT	CCGCGAAGAG	AAAGAAGAGC	GTCACCGCAC	AGCGCGGCCAG	CCCGGTCTT
	47881	TGCTCCTTCG	CACAGCGCG	GATCTGGTT	CTCCAGCAAT	TGGACCCGGA	GAGCAACGCC
	47941	TATAATCTCC	CGCTCGTCA	ACGCCGCGC	GGTCTATTGG	ACGCGCCGGC	CCTGGAGCGT
	48001	GCGCTGGCGC	TCGTCTCGC	GGCCCACGAG	GGTTGCGGA	GGTGTTCGA	CACCCCGGAC
	48061	GGCGAGCCCC	TCCAGGGGT	GCTTCCCGCC	CCGGAACACC	TCTCGGCCA	CGCGGGGGCG
	48121	GGCAGCGAGG	AGGACGGCGC	CCGGCTCGTC	CGCGACGAGA	TCGCCGCGCC	GTTCGACCTC
	48181	GCCACCGGGC	CGTTGATCG	GGCCCTGCTG	ATCCGCTCG	GTGACGACGA	CCACGTTCTC
	48241	GCGGTGACCG	TGCAACATGT	CGCCGGCGAC	GGCTGGTCGT	TCGGGCTCT	CCAACATGAA
	48301	CTCGCAGCCC	ACTACACGGC	GCTCGCGAC	ACTGCCCGCC	CTGCGAAGCT	GCCGCCGTTG
	48361	CCGGTGCAGT	ACGCCGACTT	CGCCGCCCTG	GAGCGCGCG	AACTCACCGG	CGCCGGACTG
	48421	GACAGCGTC	TGGCTACTG	GCGCGAGCAA	CTCCGGGCG	CCCCGGCGCG	GTCGCCCTC
	48481	CCCACCGACC	CTCCCCGCC	GCGGGTCGCC	GACGCGGACG	CGGGCATGGC	CGAGTGGCGG
	48541	CCGCCGGCCG	CGCTGGCCAC	CGGGTCCCTC	ACGCTCGCGC	GCGACTCCGG	TGCGTCCGTG
	48601	TTCATGACCC	TGCTGGCGC	CTTCCAAGCG	GTCCTCGGCC	GGCAGGGCGG	CACGGGGGAC
	48661	GTGCTGGTCG	GCACGCCGT	GGCGAACCGT	ACGCGGGCGG	CGTACGAGGG	CCTGATCGGC
	48721	ATGTCGTCA	ACACGCTCGC	GCTGCGCGGC	GACCTCTCG	GGGATCCGTC	GTTCCGGGAA
	48781	CTCCTCGACC	GCTGCCGGGC	CACGACCAACG	GACCGCTTCG	CCCACGCCGA	CCTGCCGTT
	48841	GAGAACGTCA	TCGAACCTGT	CGCACCGGAA	CGCGACCTGT	GGTCAACCC	GGTCGTCCAG
	48901	GTGCTGTTGC	AGGTGCTGCG	GGCGGACGCC	GCGACGGCCG	CGCTGCCCCG	CATCGCGGCC
	48961	GAACCGTTCC	GCACCGGACG	CTGGTTCAC	CGCTTCGACC	TGAATTCCA	TGTGTACGAG
	49021	GAGCCGGGTG	GGCGCGTGA	CGCGAACCTG	CTCTACAGCC	GTGCGCTGTT	CGACGAGCCA

49081 CGGATCACGG GGTTGCTGGA GGAGTTCACG GCGGTGCTTC AGGCGGTAC CGCCGACCCG
 49141 GACGTACGGC TGTCGCGGCT GCGGCGCCGGC GACCGACGG CGGCAGCGCC CGTGGTCCCC
 49201 TCGAACGACA CGGGCGGGGA CCTGCCCGTC GACACGCTGC CGGGCCTGCT GGCCCGGTAC
 49261 GCCGCACGCA CCCCCGGCGC CGTGGCGTC ACCGACCCGC ACATCTCCCT CACCTACGCG
 5 49321 CAGCTGGACC GGCAGGGCGAA CGGCCCTCGCG CACCTGCTCC GCGCGCGCGG CACGCCACC
 49381 GGCAGACCTGG TCGGGATCTG CGCCGATCGC GGCGCCGACC TGATCGTCGG CATCGTGGGG
 49441 ATCCTCAAGG CGGGCGCCGC TTATGTGCCG CTGGACCCCG AACATCTCC GGAGCGACG
 49501 GCGTTCGTGC TGGCCGACGC GCAAGCTGACC ACGGTGGTGG CGCACGAGGT CTACCGTTCC
 49561 CGGTTCCCCG ATGTGCCGCA CGTGGTGGCG TTGGACGACC CGGAGCTGGA CGGGCAGCCG
 10 49621 GACGACACGG CGCCGGACGT CGAGCTGGAC CGGGACAGCC TCGCCTACGC GATCTACACG
 49681 TCCGGGTCGA CGGGCAGGCC GAAGGCCGTG CTCATGCCGG GTGTCAAGCGC CGTCAACCTG
 49741 CTGCTCTGGC AGGAGCGCAC GATGGGCCGC GAGCCGGCCA GCGCACCGT CCAGTTCTG
 49801 ACGCCCACGT TCGACTACTC GGTGCAGGAG ATCTTTCCG CGCTGCTGGG CGGCACGCTC
 49861 GTCATCCCACGG CGGACGAGGT GCGGTTGAC CGGCCGGGAC TCGCCCGGTG GATGGACGAA
 15 49921 CAGGCAGATTA CCCGGATCTA CGCGCCGACG GCGTACTGC GCGCGCTGAT CGAGCACGTC
 49981 GATCCGCACA GCGACCAGCT CGCCGCCCTG CGGCACCTGT GCCAGGGCGG CGAGGCGCTG
 50041 ATCCTCGACG CGCGGTTGCG CGAGCTGTGC CGGCACCGGC CCCACCTGCG CGTGACAAT
 50101 CACTACGGTC CGGCCGAAAG CGAGCTCATC ACCGGTACA CGCTGCCCGC CGACCCCGAC
 50161 GCGTGGCCCG CCACCGCACC GATCGGCCCG CGATCGACA ACACCCGCAT CCATCTGCTC
 20 50221 GACGAGGCAGA TGCGGCCGGT TCCGGACGGT ATGCCGGGC AGCTCTGCGT CGCCGGCGTC
 50281 GGCTCTGCCCG GTGGGTACCT GGGCCGTCCC GAGCTGACCG CCGAGCGCTG GGTGCCGGGA
 50341 GATCCGGTGC CGGAGGAGCG CATGTACCTC ACCGGCGACC TGGCCCGCCG CGGCCCGAC
 50401 GGCAGACCTGG AATTCTCGG CGCGATCGAC GACCAGGTCA AGATCCGGG CATCCCGCTC
 50461 GAACCGGGGTG AGATCGAGAG CCTGCTCGC GAGGACGCC GCGTCACGCA GGCAGCGGTG
 25 50521 TCCGTGCGCG AGGACCGGCG GGGCGAGAAG TTCTGGCCG CGTACGTCGT ACCGGTGGCC
 50581 GCCGGCACG GCGACGACTT CGCCGCGTGC CTGCGCGCG GACTGGCCG CGGGCTGCC
 50641 GCCCGCTCG TGCCCTCCGC CGTCGCTCTG GTGGAGCGAC TGCCGAGGAC CACGAGCGGC
 50701 AAGGTGGACC GGCAGCGCTG GCCCCGACCCG GAGCCGGGCC CGGCCTCGAC CGGGGGGTT
 30 50761 ACGCCCCGCA CGATGCCGA GCGGACGGTG TGCGGATCT TCCAGGAGGT GCTCAGCTC
 50821 CGCGGGGTCG GTGCCGACGA CGACTTCTTC ACGCTGGCG GGCACCTCCCT GCTCGCACCC
 50881 CGGGTGTCT CCCGCATCCG CGCCGAGCTG GTGCCGATG TCCCGCTSGC TACGCTCTC
 50941 GACGGGCGGA CGCCCGCCGC GCTCGCCCGT CGGGCGGACG AGGCGGGCCC GGCCGCCCTG
 51001 CCCCCGATCG CGCCCTCCGC GGAGAACCGG CGGCCCCCCC TCACCGCGGC ACAGGAACAG
 51061 ATGCTGACT CGCACGGCTC GCTGCTCGCC GCGCCCTCCT ACACGGTTCG CCGTACGGG
 35 51121 TTCCGGCTGC CGGGGCCACT CGACCGCGAA GCGCTCGACG CGGCACTGAC CGGGATCGCC
 51181 GCGGCCACG AGCCGCTGCG GACCGGGGTT CGCGATCGGG AACAGGTCGT CGGGCGGCC
 51241 GCTCCGGTGC CGGCCGAGGT GTTCCGGTG CGGTCGGCG ACGTCGACGC CGCGGTCCGG
 51301 GTCGCCCACC GGGAGCTGAC CGGGCCGTT GACCTCGTA ACGGGTCGTT GCTCGTGCC
 51361 GTGCTGCTGC CGCTGGCGC CGAGGATCAC GTGCTGCTGC TGATGCTGCA CCACCTCGCC
 40 51421 GGTGACGGAT GGTCTTCGA CCTCCTGGTC CGGGAGTTGT CGGGGACGCA ACCGGACCTT
 51481 CGGGTGTCT ACACGGACGT GGGCCGGTGG GAAACGGAGTC CGGCCGTGAT CGCGGCCAGG
 51541 GAGAACGACC GGGCCTACTG GCGCCGGCGG CTGGGGGGCG CCACCGCGCC GGAGCTGCC
 51601 GCGGTCCGGC CGGGCGGGGC ACCGACCGGG CGGGCGTTCC TGTGGACGCT CAAGGACACC
 51661 GCGTCCCTGG CGGCACGGCG GGTCCGGAC GCCCACGACG CGACGTTGCA CGAAACCGTG
 45 51721 CTCGGCGCCT TCGCCCTGGT CGTGGCGGAG ACCGCCGACA CGCACGACGT GTCGTCGCG
 51781 ACGCCGTTCG CGGACCGGGG GTACGCCGG ACCGACCACC TCATCGGCTT CTTCCGGAAG
 51841 GTCCCTCGC GCGCCTCGA CCTCGGGCGC ACGCCGTCGT TCCCGGAGGT GCTCGCCGG
 51901 GTGCCACACCG CGATGGGGG CGCGCACGCC CACCAAGCGG TGCCCTACTC CGCGCTGCGC
 51961 GCGGAGGACC CGCGCTGCC GCGGGCCCCC GTGCTGTTCC AGCTCATCAG CGCGCTCAGC
 50 52021 GCGGAACCTGC GGCTGCCCGG CATGCACACC GAGCCGTTCC CGCTCGTCC CGAGACCGTC
 52081 GACGAGATGA CGGGCGAACT GTCGATCAAC CTCTTCGACG ACGGTCGAC CGTCCTCCGG
 52141 GCGGTGGTCC ACGATGCCGC GTGCTCGAC CGTGCACCG TCGACGATTT GTCACCCCGG
 52201 GTGAGGGCGA CGCTCGTGC CGCCGCGGGG GACCTCACCG TACCGCTCAC CGGTTACGTG
 52261 GAAAGCGAGT AGCCATGCCG GAGCAGGACA AGACAGTCGA GTACCTTCG TGGGCAGCG
 55 52321 CGGAACCTCA GAAGACCGT GCGGAACCTCG CGCGCACAG CGAGCCGTTG GCGATCGTGG
 52381 GGATGGCCTG CGGGCTGCCG GCGGGGGTCC CGTCCGGGGA GGACCTGTGG CAGTCTGG
 52441 AGTCGGTGG CGACGGCATC ACCCGCTTCC CCACGGACCG GGGCTGGGAG ACCACCGCCG
 52501 ACGGTGCGGG CGGCTTCCCT ACCGGGGCGG CGGGCTTCCG CGCGGGCGTTC TTCGGCATCA
 52561 GCCCGCGCGA GGCCTGGCG ATGGACCCGC AGCAGCGCCT GGCCTGGAG ACCTCGTGGG
 60 52621 AGGCAGTTGCA GCACGGGGC ATCGATCCGC AGACGCTGCG GGGCAGTGCAC ACGGGGGTG
 52681 TCCTCGGCCG GTTCTTCAG GGGTACGGCA TCGCGCCGA CTTCGACGGT TACGGCACCA
 52741 CGAGCATTCA CACGAGCGTG CTCTCCGGCC GCCTCGCGTA CTTCTACGGT CTGGAGGGTC
 52801 CGGGCGTCAC GGTGCACAG GCGTGTTCGT CGTCGCTGGT GGCGCTGAC CAGGCCGGGC
 52861 AGTCGCTGCG CTCCGGGAA TGTCGCTCG CCCTGGTCGG CGCGCTCACG GTGATGGCCT

5	52921	CGCCGGCGGG	GTTCGCGGAC	TTCTCCGAGC	AGGGCGGCCT	GGCCCCCGAC	GCGCGCTGCA
	52981	AGGCCTTCGC	GGAAGCGGCT	GACGGCACCG	GTTTCGCCGA	GGGGTCCCGC	GTCTGATCG
	53041	TCGAGAAGCT	CTCCGACGCC	GAGCGAACG	GCCACCGCGT	GCTGGCGTC	GTCCGGGGTT
	53101	CCGCCGTCAA	CCAGGACGGT	GCCTCCAACG	GGCTGTCCGC	GCCGAACGGG	CCGTCGCAGG
5	53161	AGCGGGTGT	CGGGCAGGCC	CTGGCCAACG	CCGGACTCAC	CCCGGCGGAC	GTGGACGCCG
	53221	TCGAGGCCA	CGGCACCGGC	ACCAGGCTGG	GCGACCCCAT	CGAGGCACAG	GCCGTGCTGG
	53281	CCACCTACGG	GCAGGGCGC	GACACCCCTG	TGCTGCTGGG	CTCGCTGAAG	TCCAACATCG
	53341	GCCACACCCA	GGCCGCCGCG	GGCCTCGCCG	GTGTATCAA	GATGGTCCCTC	GCCATGCGGC
10	53401	ACGGCACCCCT	GCCCCGCACC	CTGCACGTGG	ACACGCCGTC	CTCGCACGTC	GACTGGACGG
	53461	CCGGCGCCGT	CGAACCTCTC	ACCGACGCC	GGCCCTGGCC	CGAAACCGAC	CGCCACCGGC
	53521	GCGCCGGTGT	CTCCTCCTTC	GGCGTCAGCG	GCACCAACGC	CCACATCATC	CTCGAAAGCC
	53581	ACCCCCGACC	GGCCCCCGAA	CCCGCCCCCG	CACCCGACAC	CGGACCGCTG	CCGCTGCTGC
15	53641	TCTCGGCCCG	CACCCCGCAG	GAACCTCGACG	CACAGGTACA	CCGCTTGCGC	GCGTCCCTCG
	53701	ACGACAACCC	CGGCGGGAC	CGGGTCGCGC	TCGCGCAGAC	ACTCGCCCCG	CGCACCCAGT
	53761	TCGAGCACCG	CGCGTGTG	CTCGCGACA	CGCTCATCAC	CGTGAGCCCG	AACGCCGGCC
	53821	GCGGACCGGT	GGTCTCGT	TACTCGGGC	AAAGCACGCT	GCACCCGCAC	ACCGGGCGGC
	53881	AACTCGCGTC	CACCTACCCC	GTGTTCGCCG	AAGCGTGGCG	CGAGGCCCTC	GACCAACCTCG
20	53941	ACCCCCACCA	GGGCCCCGGC	ACCGACTTCG	CCCACAGAC	CGCGCTCAC	GCGCTCCTGC
	54001	GGTCTGGGG	CATCACCCCC	CACGCGTCA	TCGGCCACTC	CTCGGTGAG	ATCACCGCCG
	54061	CGCACGCCG	CGGTGTCTG	TCCCTGAGGG	ACGCGGGCGC	GCTCCTCAC	ACCCGCACCC
	54121	GCCTGATGGA	CCAACTGCCG	TCGGCGGGC	CGATGGTCAC	CGTCCTGACC	AGCGAGGAAA
	54181	AGGCACGCCA	GGTGTGCGG	CGGGCGTGG	AGATCGCCG	CGTCAACGGC	CCCCACTCCC
	54241	TCGTGCTGTC	CGGGGACGAG	GAAGCGTAC	TCGAAGCCGC	CGGCGAGCTC	GGCATCCACC
25	54301	ACCGCCTGCC	GACCCGCCAC	GCCGGCCACT	CCGAGCGCAT	GCAGCCACTC	GTCGCCCCCCC
	54361	TCCCTGACGT	CGCCCCGACC	CTGACGTACC	ACCAGCCCCA	CACCGCCATC	CCCGGCGACC
	54421	CCACCAACCGC	CGAACATACTGG	GCGCACCAAGG	TCCCGGACCA	AGTACGTTTC	CAGGGCGACA
	54481	CCGAGCAGTA	CCCGGCGCG	ACGTTCTCG	AGATCGCC	CAACCAGGAC	CTCTCGCCGC
	54541	TCGTCGACGG	CGTTGCCGCC	CAGACCGGT	CGCCCGACGA	GGTGCGGGGC	CTGCACACCG
30	54601	CGCTCGCGCA	GCTCCACGTC	CGCGCGTGC	CGATCGACTG	GACGCTCGTC	CTCGCGGGGG
	54661	ACCGCGCGCC	CGTCACGTC	CCCACGTATC	CGTTCCAGCA	CAAGGACTAC	TGGCTGCGGC
	54721	CCACCTCCCG	GGCCGATGTG	ACCGGGCGGG	GGCAGGAGCA	GGTGGCGCAC	CCGCTGCTCG
	54781	GCGCCGCGGT	CGCGCTGCC	GGCACGGGC	GAGTCGTCT	GACCGGGCGC	CTGTCGCTGG
	54841	CCTCCCATCC	GTGGCTCGGC	GAGCACGCGG	TCGACGGCAC	CGTGCCTCTG	CCCGGCGCGG
35	54901	CCTTCCTCGA	ACTCGCGCG	CGCGCCGGCG	ACGAGGTGCG	CTGCGACCTG	CTGCACGAAC
	54961	TCGTCATCGA	GACGCCGTC	GTGCTGCCCG	CGACCGGGCG	TGTGGCGGTC	TCCGTCGAGA
	55021	TCGCCGAACC	CGACGACACG	GGCGGGCGGG	CGGTACCCGT	CCACCGCGG	GCCGACGGCT
	55081	CGGGCCTGTG	GACCCGACAC	GCCGGCGGAT	TCCTCGGCAC	GGCACCGGCA	CCGGCCACGG
	55141	CCACGGACCC	GGCACCCCTGG	CCGCCCCGCG	AAGCGGGACC	GGTCGACGTC	GCCGACGTCT
40	55201	ACGACCGGTT	CGAGGACATC	GGGTACTCTC	ACGGACCGGG	CTTCCGGGGG	CTGCGGCGCUG
	55261	CCTGGCGCGC	CGGCGACACC	GTGTACGCCG	AGGTGCGCT	CCCCGACGAG	CAGAGCGCCG
	55321	ACGCCGCCCC	TTTCACGTC	CACCCCGCGC	TGCTCGACGC	CGCGTTCCAG	GCCGGCGCGC
	55381	TGGCCGCGCT	CGACGCACCC	GGCGGGGGCGG	CCCGACTGCC	GTTCTCGTTC	CAGGACGTCC
	55441	GCATCCACGC	GGCGGGGGCG	ACCGGGCTGC	GGGTACCGGT	CGGCGCGAC	GGCGAGCGCA
45	55501	GCACCGTCCG	CATGACCGGC	CCGGACGGGC	AGCTGGTGGC	CGTGGTCGGT	GCCGTGCTGT
	55561	CGCGCCCGTA	CGCGGAAGGC	TCCGGTGACG	GCCTGCTGC	CCCGGTCCTGG	ACCGAGCTGC
	55621	CGATGCCCGT	CCCGTCCGCC	GACGATCCGC	GGCTGGAGGT	CCTCGCGGCC	GACCCGGGGCG
	55681	ACGGCGACGT	TCCGGCGGCC	ACCGGGGAGC	TGACCCCGCC	CTTCCCTCGGC	GGCGTCCAGC
	55741	GCCACCTGTC	CGCGCCCGAG	GACACCCACT	TGGTGGTACG	GACCGGGCAC	GGCCCGGGCG
50	55801	CTGCCGCCG	CGCGGGTCTG	GTCCGCTCG	CGCAGGGCGA	GAACCCCGGC	CGCGTCGTGC
	55861	TCGTCGAGGC	GTCCCCGGAC	ACCTCGGTGG	AGCTGCTCGC	CGCGTGC	GCGCTGGAGC
	55921	AACCGCAGCT	GGCGTCCGG	GACGGCGTC	TCTTCGCGCC	GGGCGTGT	CGGATGTCGG
	55981	ACCCCGCGCA	CGGCCCCGCTG	TCCCTGCCGG	ACGGCGACTG	GCTGCTCAC	CGGTCCGCGCT
	56041	CGGGCACGTT	GCACGACGTC	GCGCTCATAG	CCGACGACAC	GCCCCGGCGG	GCGCTCGAAG
55	56101	CGGGCGAGGT	CCGCATCGAC	GTCCGCGCG	CCGGACTGAA	CTTCCCGCGAT	GTGCTGATCG
	56161	CGCTCGGGAC	GTACACCGGG	GCCACGGCCA	TGGCGGGCGA	GGCCGCGGGC	GTCTGGTGG
	56221	AGACCGGGCC	CGGCGTGGAC	GACCTGTCCC	CCGGCGACCG	GGTGTTCGGC	CTGACCCGGG
	56281	GCGGCATCGG	CCCGACGGCC	GTCACCGACC	GGCGCTGGCT	GGCCCGGATC	CCCGACGGCT
	56341	GGAGCTTCAC	CACGGCGGCC	TCCGTCCCGA	TCGTGTTCGC	GACCGCGTGG	TACGGCCCTGG
60	56401	TCGACCTCGG	CACACTGCC	GCCGGCGAGA	AGGTCTCGT	CCACCGGGCC	ACCGGGCGTG
	56461	TCGGCATTGGC	CGCCGACAG	ATCGCCCGCC	ACCTGGCGC	CGAGCTCTAC	GCCACCGCCA
	56521	GTACCGGCAA	GCAGCACGTC	CTGCGCGCC	CCGGCGTGC	CGACACGAC	ATCGCCGACT
	56581	CTCGGACGAC	CGCGTCCCG	ACCGCTTTCC	CGCGCATGGA	CGTGCCTCTG	AACCGCCTGA
	56641	CGGGCGAGTT	CATCGACGCC	TCGCTCGACC	TGCTGGACGC	CGACGGCCGG	TTCGTCGAGA
	56701	TGGGCCGCAC	CGAGCTGCC	GACCCGGCCG	CGATGTC	CGCCTACCTG	CCGTTGACCC

56761	TGCTGGACGC	GGGCGCCGAC	CGCATCGCG	AGATCCTGGG	CGAACTGCTC	CGGCTGTTGC	
56821	ACGCCGGCGC	GCTGGAGCCG	CTGCCGGTCC	GTGCCTGGGA	CGTCGGCAG	GCACGCCAG	
56881	CGCTCGGCTG	GATGAGCCGC	GCCC GCCACA	TCGGCAAGAA	CGTCCTGACG	CTGCCCCGGC	
56941	CGCTCGACCC	GGAGGGCGCC	GTCGTCCCTCA	CCGGCGGCTC	CGGCACGCTC	GCCGGCATCC	
5	57001	TCGCCCCGCCA	CCTGCGCGAA	CGGCATGCT	ACCTGCTGTC	CCGGACGGCA	CGGCCCCGAGG
57061	GGACGCCCGG	CGTCCACCTG	CCCTGCGACG	TCGGTGACCG	GGACCAGCTG	CGGGCGGGCC	
57121	TGGAGCGGGT	GGACCGGCCG	ATCACCGCCG	TGGTGACCT	CGCCGGTGC	CTGGACGACG	
57181	GCACCGTCGC	GTCGCTCACC	CCCGAGCGTT	TCGACACGGT	GCTGCGCCCG	AAGGCCGACG	
57241	GCGCCTGGTA	CCTGACAGAG	CTGACGAAGG	AGCAGGACCT	CGCCGCGTTC	GTGCTCTACT	
10	57301	CGTCGGCCGC	CGGCGTGCTC	GGCAACGCG	GCCAGGGCAA	CTACGTGCCC	GCAGAACGCGT
57361	TCCTCGACGC	GCTCGCCGAG	CTGCGCCACG	GTTCCGGGCT	GCCGGCCCTC	TCCATCGCT	
57421	GGGGGCTCTG	GGAGGACGTG	AGCGGGCTCA	CCGCGGCGCT	CGGCGAAGCC	GACCGGGACC	
57481	GGATGCGGGC	CAGCGGTTTC	CGGGCCATCA	CCGCGCAACA	GGGCATGCAC	CTGTACGAGG	
15	57541	CGGCCGGCCG	CACCGGAAGT	CCCGTGGTGG	TCGCGCGGC	GCTCGACGAC	GCGCCGGACG
57601	TGCCCGCTGCT	GC CGGCCCTG	CGGCGGACGA	CCGTCGGCGC	GGCCGCCGTC	CGGGAGTGT	
57661	CGTCCGCCGA	CCGGCTCGCC	GGCGTGACCG	GCGACGAGCT	CGCCGAAGCG	CTGCTGACGC	
57721	TCGTCGGGA	GAGCACCGCC	GCCGTGCTCG	GCCACGTGGG	TGGCGAGGAC	ATCCCCGCGA	
57781	CGCGGGCGTT	CAAGGACCTC	GGCATCGACT	CGCTCACCGC	GGTCCAGCTG	CGAACCGCCC	
20	57841	TCACCGAGGC	GACCGGTGTG	CGGCTGAACG	CCACGGCGGT	CTTCGACTTC	CCGACCCCCG
57901	ACGTGCTCGC	CGGGAAGCTC	GGCGACGAAC	TGACCGGCAC	CCGCGCGCCC	GTGCGCCCC	
57961	GGACCGCGGC	CACGGCCGGT	GCGCACGACG	AGCCGCTGGC	GATCGTGGGA	ATGGCCTGUC	
58021	GGCTGCCCGG	CGGGGTGCG	TCACCCGAGG	AGCTGTGGCA	CCTCGTGGCA	TCCGGCACCG	
58081	ACGCCATCAC	GGAGTCCCG	ACGGACCGCG	GCTGGGACGT	CGACGCGATC	TACGACCCGG	
58141	ACCCCGACGC	GATCGCAAG	ACCTTCGTC	GGCACGGTGG	CTTCCTCACC	GGCCGCACAG	
25	58201	GCTTCGACGC	GGCGTTCTTC	GGCATCAGCC	CGCGCGAGGC	CCTCGCGATG	GACCCGCAGC
58261	AGCGGGTGTG	CCTGGAGACG	TCGTGGGAGG	CGTTCGAAAG	CGCCGGCATC	ACCCCGGACT	
58321	CGACCCGCCG	CAGCACACC	GGCGTGTTCG	TCGGCGCCTT	CTCCTACGGT	TACGGCACCG	
58381	GTGCGGACAC	CGACGGCTTC	GGCGCGACCG	GCTCGCAGAC	CAGTGTGCTC	TCEGGCCGGC	
30	58441	TGTCGACTT	CTACGGTCTG	GAGGGTCCGG	CGGTACCGGT	CGACACGGCG	TGTCGTCGT
58501	CGCTGGTGGC	GTCGACACC	GGCGGGCAGT	CGCTCGCCTC	GGCGAATGC	TCGCTCGCCC	
58561	TGGTCGGCGG	CGTCACGGT	ATGGCGTC	CCGGCGCCTT	CGTGGAGTTC	TCCCCGACG	
58621	GCGGCCTCGC	GCGGACGGC	CGGGCGAAGG	CGTTCGGCGC	GGGTGCGGAC	GGCACGAGCT	
58681	TCGCGGAGGG	TGCGGGTGTG	CTGATCGTC	AGAGGCTCTC	CGACGCCGAA	CGAACCGTC	
35	58741	ACACCGTCCT	GGCGGTGTC	CGTGGTCCGG	CGGTCAACCA	GGATGGTGCC	TCCAACGGGC
58801	TGTCGGCGCC	GAACGGCCG	TCGCAAGGAGC	GGGTGATCCG	CGAGGCCCCG	GCCAACCGCC	
58861	GGCTCACCCC	GGCGGACGTG	GACGCCGTC	AGGCCCACGG	CACCGGCACC	AGGCTGGCG	
58921	ACCCCATCGA	GGCACAGCG	GTACTGGCCA	CCTACGGACA	GGAGCGCGCC	ACCCCCCTGC	
58981	TGCTGGGCTC	GCTGAAGTCC	AACATCGGCC	ACGCCAGGC	CGCGTCCGGC	GTGCCCGCA	
40	59041	TCATCAAGAT	GGTGCAAGGC	CTCCGGCACG	GGGAGCTGCC	GGCGACGCTG	CACGCCGACG
59101	AGCCGTGCCC	GCACGTGAC	TGGACGGCCG	GGGCCGTGCA	ACTGCTGACG	TCGGCCGGC	
59161	CGTGGCCCGA	GACCGACCGG	CCACGGCGT	CCGCCGTCTC	CTCGTTCGGG	GTGAGCGGC	
59221	CCAACGCCCA	CGTCATCCTG	GAGGCCGGAC	CGGTAACCGG	GACGCCCGCG	GCATCGCCTT	
59281	CCGGTGACCT	TCCCCGTCTG	GTGTGGCAC	GTCACCCGG	AGCGCTCGAC	GAGCAGATCC	
45	59341	GCCGACTGCG	CGCCTACCTG	GACACCACCC	CGGACGTGCA	CGGGTGGCC	GTGGCACAGA
59401	CGCTGGCCCG	GCGCACACAC	TTCGCCCACC	GGCCCGTGCT	GTCGGTGAC	ACCGTCATCA	
59461	CCACACCCCC	CGCGGACCGG	CCCAGCAAC	TCGTCTTCGT	CTACTCCGGC	CAGGGCACCC	
59521	AGCATCCCGC	GATGGCGAG	CAGCTCGCC	CCGCCCCATCC	CGTGTTCGCC	GACGCCCTGGC	
59581	ATGAAGCGCT	CGGCCCTT	GACAACCCCC	ACCCCCACGA	CCCCACGCAC	AGCCAGCATG	
50	59641	TGCTTTCGC	CCACCAGCG	CGTTTACCG	CCCTCCTGCG	GTCCCTGGGC	ATCACCCCCGC
59701	ACCGGGTCAT	CGGCCACTCG	CTGGGCGAGA	TCACCGCGC	GCACGCCGCC	GGCATCCTGT	
59761	CGCTGGACGA	CGCGTGCACC	CTGATCACCA	CGCGCGCCCG	CCTCATGCAC	ACGCTCCCGC	
59821	CACCCGGTGC	CATGGTCACC	GTACTGACCA	CGAAGAGAA	GGCACGCCAG	GGGTGCGGC	
59881	CGGGCGTGG	GATCGCCGCC	GTCAACGGGC	CCCACCTCCAT	CGTGTCTGCC	GGGGACGAGG	
59941	ACGCCGTGCT	CACCGTCGCC	GGGCAGCTCG	GCATCCACCA	CCGCCTGCC	GCCCCGCACG	
55	60001	CCGGGCACTC	CGCGCACATG	GAGCCCGTGG	CCGCCAGAGCT	GTCGCCACC	ACCCGGGGC
60061	TCCGCTACCA	CCCTCCCCAC	ACCTCCATTC	CGAACGACCC	CACCAACCGCT	GAGTACTGGG	
60121	CCGAGCAGGT	CCGCAAGCCC	CTGCTGTTCC	ACGCCACGC	CGAGCAGTAC	CCGGACGCCG	
60181	TGTCGTTGG	GATCGCCGCC	GCCCAGGAC	TCTCCCCGCT	CGTCGACGGG	ATCCCCTGTC	
60241	AGAACGGCAC	CGCGGACGAG	GTGCAACGCC	TGCACACCGC	GTCGCCGAC	CTCTACGCC	
60	60301	CGGGTGCCAC	GTCGACTGG	CCCCGACATC	TCGGGGCTGG	GTACGGGCAC	GACCGGGATG
60361	TGCCCCGCGTA	CGCGTCCAA	CGGCAGGACT	ACTGGATCGA	GTCGGCACGC	CCGGCCGCAT	
60421	CCGACGCGGG	CCACCCCCGTG	CTGGGCTCCG	GTATCGCCCT	CGCCGGGTGC	CCGGGCCGGG	
60481	TGTTCACGGG	TTCCGTGCCG	ACCGGTGCGG	ACCGCGCGGT	GTTCGTCCGCC	GAGCTGGCGC	
60541	TGGCCGCCGC	GGACGCGGTC	GACTGCGCCA	CGGTCGAGCG	GCTCGACATC	GCCTCCGTGC	

60601 CCGGCCGGCC GGGCCATGGC CGGACGACCG TACAGACCTG GGTGACGGAG CGGCCGGACG
 60661 ACGGCCGGCG CCGGTTCAACC GTGCACACCC GCACCGCGA CGCCCGTGG ACGCTGCACG
 60721 CCGAGGGGGT GCTGCCCGCC CATGGCACGG CCCTGCCGA TGCGGCCGAC GCGAGGGC
 60781 CCCCACCGG CGCGGTGCCG CGGGACGGGC TGCGGGGTGT GTGGGCCGG GGGGACCAAGG
 5 60841 TCTTCGCCGA GGCGAGGTG GACGGACCGG ACGGTTCTGT GGTGACCCCC GACCTGCTCG
 60901 ACGCGGTCTT CTCCGCCGTC GGCACGGAA GCCGCCAGCC GGCGGATGG CGCGACCTGA
 60961 CGGTGCACGC GTCGGACGCC ACCGTACTGC GCGCTGCCT CACCCGGCGC ACCGACGGAG
 61021 CCATGGGATT CGCCGCCCTC GACGGCGCCG GCCTGCCGGT ACTCACCGCG GAGGCCGTGA
 61081 CGCTGCCGGA GGTGGCTCA CGCTCCGGCT CCGAGGAGTC GGACGGCTG CACCGGTTGG
 10 61141 AGTGGCTCGC GGTGCCGAG GCGGTCTACG ACGGTGACCT GCGGAGGGG CATGCTCTGA
 61201 TCACCGCCGC CCACCCGAC GACCCCGAGG ACATACCCAC CGCGGCCAAC ACCCGCGCCA
 61261 CCCCGTCTCT GACCGCCCTC CAACACCAC TCACCAACAC CGACCACACC CTACATCGTCC
 61321 ACACCAACAC CGACCCCGCC GGCGCCACCG TCACCGGCCT CACCCGCACC GCGCAGAACG
 61381 AACACCCCCA CCGCATCCGC CTCATGAAA CCGACCAACCC CCACACCCCC CTCCCCCTGG
 15 61441 CCCAACTCGC CACCCCTGAC CACCCCCAAC TCCGCTCAC CCACCCACACC CTCCACCAAC
 61501 CCCACCTCAC CCCCCCTCCAC ACCACCACCC CACCCACCAC CACCCCCCTC AACCCCGAAC
 61561 ACGCCATCAT CATCACCGGC GGCTCCGGCA CCCTCGCCGG CATCTCGCC CGCCACCTGA
 61621 ACCACCCCCA CACCTACCTC CTCTCCGCA CCCCCACCCCC CGACGCCACC CCCGGCACCC
 61681 ACCTCCCCCTG CGACGTGGC GACCCCCACC AACTCGCCAC CACCCCTCACC CACATCCCCC
 20 61741 AACCCCCTCAC CGCCATCTTC CACACCGCCG CCACCCCTCAGA CGACGGCATC CTCCACGCC
 61801 TCACCCCCGA CCGCCTCACC ACCGTCTCC ACCCCAAAGC AAACGCCGCC TGGCACCTGC
 61861 ACCACCTCAC CCAAAACCAA CCCCCCTACCC ACTTCGTCTCT TACTTCAGC GCCGCCJCCG
 61921 TCCTCGCAG CCCCCGACAA GGAAACTACG CGGCCGCCAA CGCCCTCCCTC GACGCCCTCG
 25 61981 CCACCCACCG CCACACCTC GGCCAACCC CGACCTCCAT CGCCTGGGGC ATGTTGGCACA
 62041 CCACCAAGCAC CCTCACCGGA CAACTCGACG AGCGCCGACCG GGACCGCATC CGCCGCGGCC
 62101 GTTCTCTCCC GATCACGGAC GACGAGGGCA TGCGCTCTA CGAGGGGGCC GTCGGCTCCG
 62161 GCGAGGACTT CGTCATGGCC GCGCGATGG ACCCGGCACA GCGATGACC GGCTCCGTAC
 62221 CGCCCATCTC GAGCGGCCTG CGCAGGAGCG CGCGGCCGCTG CGCCCGTGGC GGGCAGACGT
 62281 TCGCCCAGCG GCTCGCCGAG CTGCCCAGCG CCGACCGCGG CGCGGCCGCTG ACCACCTCG
 30 62341 TCTCGGACGC CACGGCGCC GTGCTCGGCC AGCGCCGACGC CTCCGAGATC GCGCCGACCA
 62401 CGACGTTCAA GGACCTCGGC ATCGACTCGC TCACCCGAT CGAGCTCGC AACCGGCTCG
 62461 CGGAGGCGAC CGGGCTCGGG CTGAGTGCCA CGCTGGTGT CGACCCACCG ACACCTCGGG
 62521 TCCTCGCCGC CAAGCTCCGC ACCGATCTGT TCGGCACGGC CGTGCACCG CCCGCGCGGA
 62581 CGGCACGGAC CCACCAACGAC GAGCCACTCG CGATCGCTGG CATGGCGTGC CGACTGCCG
 35 62641 GCGGGGTCGC CTCGCCGGAG GACCTGTGGC AGCTCGTGGC GTCCGGCACC GACCGATCA
 62701 CCGAGTTCCC CACCGACCGC GGCTGGACA TCGACCGGCT GTTCGACCCCG GACCCGGACG
 62761 CCCCCGGCAA GACCTACGTC CGGCACGGCG GCTTCTCGC CGAGGCCGCC GGCTCGATG
 62821 CGCGTTCTT CGGCATCAGC CGCGCGAGG CACGGGCCAT GGACCCGCAG CAGCGCTCA
 40 62881 TCCTCGAAAC CTCCTGGGAG GCGTTCGAGA ACGCGGGCAT CGTGCACGGAC ACGCTCGCG
 62941 GCAGCGACAC CGGCGTGTTC ATGGGCGCT TCTCCATGG TACGGCGCC GGCCTCGAC
 63001 TGGCGGGTT CGGCGCCACC GCCACGCAGA ACAGCGTGT CTCCGGCGG TTGTCGACT
 63061 TCTTCGGCAT GGAGGGCCCG GCCGTACCGC TCGACACCAC CTGCTCGTGC TCGCTGGTGC
 63121 CCCTGCACCA GGCAGCACAG GCGCTCGGGA CTGGAGAATG CTCGCTGGCG CTCGCCGGCG
 63181 GTGTACCGGT GATGCCACC CGCGTGGCT ACGTGAGTT CTGCCGCCAG CGGGGACTCG
 45 63241 CCCCCGACGG CCGTTGCCAG GCCTCGCGG AAGGCCCGA CGGCACGAGC TTCTCGGAGG
 63301 GCGCCGGCGT TCTTGTGCTG GAGCGGCTCT CCGACGCCGA GCGAACCGGA CACACCGTCC
 63361 TCGCGGTCTG CCGCTCTCC GCCGTCAACC AGGACGGCGC CTCCAACGGC ATCTCCGCAC
 63421 CCAACGGCCC CTCCCAGCAG CGCGTCATCC GCCAGGCCCT CGACAAGGGC GGGCTCGCCC
 63481 CGGCCGACGT GGACGTGGT GAGGCCACCG GCACCGGAAC CCCGCTGGGC GACCCGATCG
 50 63541 AGGCACAGGC CATCACCGCG ACCTACGGCC AGGACCGCGA CACACCGCTC TACCTCGGTT
 63601 CGGTCAAGTC GAACATCGGA CACACCGAGA CCACCGCCGG TGTCGCCGGC GTCATCAAGA
 63661 TGGTCATGGC GATGCGCCAC GGCATCGCGC CGAACGACACT GCACGTGGAC GAGCCGTGCTG
 63721 CGCATGTGGA CTGGACCGAG GTGCGGTGG AACTGCTCAC CGAGGCCAGG CGCTGGCCCG
 63781 ACGCGGGACG CCCGCGCCGC GCGGGCGTGT CGTCGCTCGG TATCAGCGGT ACGAACGCC
 55 63841 ACGTGATCCT TGAGGGTGT CCCGGGCCGT CGCGTGTGGA CGCGTCTGTT GACGGGTTGG
 63901 TGCCGTTGCC GGTGCGGCT CGGAGTGAGG CGAGTCTGCG GGGCGAGGTG GAGCCGCTGG
 63961 AGGGGTATCT GCGCGGGAGT GTGGATGTGG CGCGCGTCGC CGAGGGGTG TGCGCTGAGC
 64021 GTGCTGTCTT CGGTACCGCT CGGGTACTGC TGGGTGATGC CGGGGTGATG GGTGTGGCG
 64081 TGGATCAGCC CGTACGGTG TTGCTTTT CCGGGCAGGG TGCTCAGTGG GTGGGATGG
 60 64141 GTGTGGAGTT GATGGACCGT TCTCGGGTGT TCGCGGCTCG TATGGAGGAG TGTGCGGG
 64201 CGTGTGGCC GCACACGGGC TGGGATGTGC GGGAGATGTT GGCACGCCGG GATGTGGCG
 64261 AGCCGGTGGG GGTGGTCCAG CCGGCCAGCT GGGCGGTGCG GGTCACTCG GCCGCACTGT
 64321 GGCAGGCCCA CGGGGTGCGA CCCGACCGCG TGATCGACCA CTCCCAGGGC GAGATCGCG
 64381 CGGCCTGCGT GGCGGGGCC CTCAGCCTTG AGGACGCCGC CGCGTGGTG GCCTGCGCA

64441	GCCAGGTCAT	CGCGGGCGGA	CTGGCCGGGC	GGGGAGCGAT	GGCTTCGGTG	GCATGCCGG
64501	CCGGTGAGGT	CGGTCTGGTC	GAGGGCGTGT	GGATCGCGC	GCGTAACGGC	CCCCCCTCGA
64561	CAGTCGTGGC	CGGCAGGCCG	TCGGCGGTGG	AGGACGTGGT	GACCGGGTAT	GAGACCGAAG
64621	CCGTGCGAGT	GCGTCGTATC	GCCGTCGACT	ACGCCTCCCA	CACGCCAAC	GTGAAACCCA
5	64681	TCGAGGACGA	ACTCGCTGAG	GTACTGAAGG	GAGTGCAGG	GAAGGCCCG
	64741	GGTGGTCGAC	CGTGGACAGC	GCCTGGGTGA	CCGAGGCCGGT	GGATGAGAGT
	64801	GGAACCTGCG	TCGCCCCGTC	GCGCTGGACG	CGGCGGTGGC	GGAGCTGGAC
	64861	TCGTGGAGTG	CAGCGCCCAT	CCGGTGTGTC	TGCGGGCGAT	GGAACAGGGC
	64921	CGTCGTTGCC	CACCGGTGAC	GGCGGCTGGG	AGCGATGGCT	GACGGCGTTG
10	64981	GGACCCCTGGG	CGCGGCAGTG	GACTGGGACA	CGGTGGTCGA	ACCGGGTCCA
	65041	TCGATCTGCC	CACCTACCGC	TTCGAGCGCC	GGCGCTACTG	GCTGGAAGCG
	65101	CCGACCTGTC	CGCGGCCGGG	CTGACAGGGG	CAGCACATCC	CATGCTGGCC
	65161	CACTACCCGC	CGACGACGGT	GGTGTGTTTC	TCACCGGCCG	GATCTCGTTG
	65221	CCTGGCTGGC	TGATCACCGC	GTGCGGGGCA	CGGTCTGCT	GCCGGGACAG
15	65281	AGCTGGTCAT	CCGGGCCGGT	GACGAGACCG	GTGCGGGAT	AGTGGATGAA
	65341	AATCCCCCCT	CGTGGTGCGC	GCGACCGCAG	CCGTGGATCT	GTCGGTGACC
	65401	CTGACGAGGC	CGGACGGCGG	CGAGTGACCG	TCCACGCCCG	CACCGAAGGC
	65461	GGACCCGGCA	CGCCAGCGGC	ACCCCTGACCC	CCGACACCCC	CGACACCCCC
	65521	GTGTTGTCGG	TGCGGAGCCG	TTCTCGCAGT	GGCCACCTGC	CACTGCCCG
20	65581	CCTCGGAGTT	CTACTTGCAC	CTGGACGCGC	TGGGCTACCG	GTTCGGACCC
	65641	GAATGCGGGC	TGCGCTGGCGT	GATGGTGACA	CCGTGTACGC	CGAGGTCGCG
	65701	ACCGTGCCGC	CGACGCGGAC	GGTTTCCGGCA	TGCACCCGGC	GTCGCTCGAC
	65761	AGAGCGGCAG	CCTGCTCATG	CTGGAATCGG	ACGGCGAGCA	GAGCGTGCAA
	65821	CCTGGCACCG	CGTCCCGTTC	CACCGCAGCG	GCGCAGCAT	GTCGCGTAC
25	65881	CGGGCCCGGA	CGGCCTCCGG	CTGCATGCCG	CGGACAGC	GAACCGTCCC
	65941	TCGACGCGCT	CGTGACCCGG	TCCCCGGAAG	CGGACCTCGC	GCCCGCCGAT
	66001	GGGTGCGGTG	GGCCCCCGGTG	CCCGTACCTG	CCGGGGCCGG	TCCGTCCGAC
	66061	TGACGCTGCC	CGGGCAGCAC	GCCGACCCGC	TCGGGGAGAC	CCGGGACCTG
	66121	TTCTGACGC	GCTGCTCCGG	GCCGACCCGG	CGGTGATCTT	CCAGGGTACCC
30	66181	CCGCCAAGGC	GGCGCGAGGC	CTGGTCCGCA	CCGCTCAGAA	CGAGCAGCCC
	66241	TCCCTGTCGA	AACGGACCCG	GGAGAGGTCC	TGGACGGCGC	GAAGCGCGAC
	66301	CACTCGCGA	GCCCCATGTG	CGGCTGCGC	ACGGCCTCTT	CGAGGCAGCC
	66361	GGGCCACGCC	GTCCCTGACG	CTCCCGGACA	CCGGGTGCGT	GCAGCTGCGG
	66421	CGGGTTCCCT	CGACGACCTT	GCCGTCGTCC	CCACCGACGC	CCCAGGACCGG
35	66481	CCGGCGAGGT	GCGGATCGCG	GTACGCGCG	CGGGCCTGAA	CTTCCGGGAT
	66541	CGCTCGGTG	GGTCGCGGAT	GCGCGTCCGC	TCGGCAGCGA	GGCCGCGGGT
	66601	AGACCGGGCC	CGGTGTGAC	GACCTGGCGC	CCGGCGACCG	GGTCTGGGG
	66661	GCCCTTCGG	ACCGGTCGCG	ATCACCGACC	GGCGGCTGCT	CGGACGGCT
	66721	GGACGTTCCC	GCAGGGCGCG	TCCGTATGA	CCGCGTTCGC	GACCGCGTGG
40	66781	TCGACCTGGC	CGGGCTGCC	CCCAGCGAGA	AGGTCTCTGAT	CCACCGCGG
	66841	TCGGCGCGGC	GGCCGTCAG	ATCGCGCGG	ATCTGGGCGC	GGAGGTGTAC
	66901	GCGCCCGGAA	GCGCCATCTG	GTGGACCTGG	ACGGAGCGCA	TCTGGCCGAT
	66961	CCGCGTTCGC	CGACGCGTTC	CCGGCGGTGCG	ATGTCTGCT	CAAATCGCTC
	67021	TCCCTGACGC	GTCCGTGCC	CTGCTCGCG	CGGGTGGCCG	GTTCATCGAG
45	67081	CGGACATCCG	GCACGCCGTC	CAGCAGCCGT	TCGACCTGAT	GGACGCCG
	67141	TGCAGCGGAT	CATCGTCAG	CTGCTCGCG	TGTTCGCG	CGACGTGCTG
	67201	CGGTCCACGC	CTGGGACGTG	CGGCAGGCGC	GGGAGGC	GGGCTGGATG
	67261	GTCACACCGG	CAAGCTGGTG	CTGACGGTCC	CGCGGCCGCT	GGATCCCGAG
	67321	TCATCACCGG	CGGCTCCGGC	ACCCCTCGCG	GCATCTCGC	CCGCCACCC
50	67381	ACACCTACCT	GCTCTCCCGC	ACCCCCACCC	CCGACACCAC	CCCCGGCACC
	67441	GCGACGTCGG	CGACCCAC	CAACTCGCCA	CCACCCCTCGC	CCGCATCCCC
	67501	CCGCCGTCTT	CCACACCGC	GGAAACCTCGC	ACGACGCCCT	GTCGACAAAC
	67561	ACCGCGTCGA	CACCGCTCTC	AAACCCAAGG	CCGACGCCG	CTGGCACCTG
	67621	CCCGCGACAC	CGACCTCGCC	GGCTTCGTC	TCTACTCCG	GGTCGCGG
55	67681	CCCCGGGGCA	GGGCAACTAC	TCGCGGGCGA	ACGCGTTCT	CGACGCGCTC
	67741	CCCCGTGCCA	AGGGCTGCC	GGCGAGTCCC	TCGCATGGGG	CATGTGGGG
	67801	CGCTCACCGC	GAAACTCACC	GACCGGGACC	GCCAGCGCAT	CCGGCGCAGC
	67861	CGTTGAGCGC	CGCGGACGGC	ATGCGGCTGT	TCGACGCGC	GACCGCTAAC
	67921	TCGTCGTCG	GACGACCGTC	GACCTCACCC	AGCTCGACGG	CGCCGTCGCG
60	67981	CGGGTCTGGC	CGCGCACCGG	GCCGGGCCG	CGCGCACGGT	CGCCCGCAAC
	68041	AGCCCCCTGGC	CGTGGCTCTT	GCCGGGCCG	CCGCGGCCG	GCAGCGGCC
	68101	AGGTCGTGCT	CCGCCACGCG	GCCGCGGTCC	TCGCGTACGG	GCTGGGCGAC
	68161	CGGACCGTCC	GTTCCGCGAG	CTCGGTTTCG	ATTGCGTAC	CGCGGTCGAC
	68221	GGCTCGCGC	CGAGACGGGG	CTGCGGCTGC	CGACGACGCT	GGTGTTCAGC

68281	CGGAGGCGCT	CACCGCCCAC	CTGCTCGACC	TGATCGACGC	TCCCACCGCC	CGGATCGCCG
68341	GGGAGTCCCT	GCCC CGGGTG	ACGGCCGCTC	CCGTGGCGGC	CGCGCGGGAC	CAGGACGAGC
68401	CGATCGCCAT	CGTGGCGATG	GGGTGCGCGG	TGCCCCGTGG	TGTGACGTCG	CCCGAGGACC
68461	TGTGGCGGCT	CGTCGAGTCC	GGCACCGACG	CGATCACCAC	GCCTCCTGAC	GACCGCGGCT
5	68521	GGGACGTCGA	CGCGCTGTAC	GACGCGGACC	CGGACCGGGC	CGGCAAGGCG
	68581	GGGGCGGITA	CCTGGCGGGG	GGGGCGGAGT	TCGACCGGGC	TTCTTCGAC
	68641	GCGAAGCGCT	CGGCATGGAC	CCGCAGCAAC	GCCTGCTGCT	CGAAACCGGC
	68701	TCGAGCGCGG	CCGGATCAGT	CCGGCGTCGC	TCCGCGGCCG	GGAGGTCGGC
10	68761	GTGCGGCCGC	GCAGGGCTAC	GGGCTGGGCG	CCGAGGACAC	CGAGGGCCAC
	68821	GTGTTTCCAC	GAGCCTGCTG	TCCGGACGGC	TGGCTACGT	GCTCGGGCTG
	68881	CGGTACCGT	GGACACGGG	TGCTCGTCGT	CTCTGGTCG	GCTGCATCTG
	68941	GGCTGCGCCT	GGCGAGTGC	GAACTCGCTC	TGGCCGGAGG	GGTCTCCGTA
15	69001	CGGCCGCGTT	CGTGGAGTTC	TCCC GCCAGC	GCAGGGCTCG	GGCCGACGGG
	69061	CGTCTGGCGC	GGGCCGGAC	GGCACGACGT	GGTCCGAGGG	CTGGGGCGTG
	69121	AACGGCTCTC	CGACGCCGAG	CGGCTCGGGC	ACACCGTGT	CGCGCGAGCG
	69181	CCGTACGTC	CGACGGCGCC	TCCAACGGCC	TCACCGCGCC	GAACCGGEC
	69241	GGGTCATCCG	GAAGGGCGTC	GCCGCGGCCG	GGCTGACCGG	CGCCGACGTG
20	69301	AGGGGCACGG	CACCGGCACC	CGGCTCGGCC	ACCCGGTCCA	GGCCGACGGG
	69361	CGTACGGGCA	GGACCGTCCG	GCACCGGTCT	GGCTGGCTC	GCTGAAGTCG
	69421	ATGCCACGGC	CGCGGCCGGT	GTCGCGGGCG	TCATCAAGAT	GGTGAGGGCG
	69481	GCACGATGCC	CGGGACGCTG	CATGTGGAGG	AGCCCTCGCC	CGCCGTCGAC
25	69541	GACAGGTGTC	CTTGCTCGGC	TCCAACCGGC	CCTGGCCGGA	CGACGAGCGT
	69601	CGGCCGTCTC	CGCGTTCGGG	CTCAGCGGGA	CGAACCGC	CGTCATCTG
	69661	GTCCGGCGCC	CGTGGCGTCC	CAGCCGCCCC	GGCCGCCCCG	TGAGGAGTCC
	69721	CGTGGGTGCT	CTCCGGCGGG	ACTCCGGCCG	CGCTGCGGGC	CCAGGGCGCC
	69781	ACCACCTCGC	GGCGGCACCG	GACGCGGATC	CGTTGGACAT	CGGGTACGGC
	69841	GCCGCGCCCA	GTTCGCCCCAC	CGTGC CGCGG	TCGTCGCCAC	CACCCGGAC
	69901	CCCGCGCTGA	CGGCCTCGCG	GACGGCGCGG	AGGCGCCCCG	AGTCGTACC
30	69961	AGGAGCGGCG	CGTCGCCTTC	CTTTCGACG	GCCAGGGCGC	CCAGCGCGCC
	70021	GCGAGCTCCA	CCGCCGGTT	CCCGTCTTCG	CGGCCGCGTG	GGACGAGGTG
	70081	TCGGCAAGCA	CCTCAAGCAC	TCCCCCACGG	ACGTCTACCA	CGCGAACAC
	70141	CCCATGACAC	CCTGTACGCC	CAGGCCGGCC	TGTTCACGCT	CGAACGAGCG
	70201	TGCTGGAGCA	CTGGGGGTG	CGGCCGGACG	TGCTCGTCGG	GCACTCCGTC
35	70261	CCGCGCGTGA	CGCGCGGGGG	GTGCTCACCC	TGGCGACGC	GACGGAGTTG
	70321	GGGGCGGGC	GCTGC CGGGCG	CTGCCGCCCCG	GGCGATGCT	CGCCGTCGAC
	70381	CGGAGGT CGG	CGCCCGCACG	GATCTGGACA	TCGCCGCGGT	CAACGGCCCC
	70441	TGCTCGCCGG	TTCGCCGGAC	GATGTGGCGG	CGTTCGAACG	GGAGTGGTCG
	70501	GGCGCACGAA	ACGGCTCGAC	GTCGGGCACG	CGTTCCTACT	CGGGCACGTC
40	70561	TCGACGGCTT	CCGTACGGTG	CTGGAGTCGC	TCGCCGTCGG	CGCGCGCGGG
	70621	TGTCCACGAC	GACGGGCCGG	GACGCCGCGG	ACGACCTCAT	AACGCCCGCG
	70681	GCCATGCGCG	TCGGCCGGTG	CTGTTCTCGG	ATGCCGTCGG	GGAGCTGGCC
	70741	TCACACCGTT	CGTGGCGTGC	GGCCCCCTCGG	GCTCCCTGGC	GTCGGCCGCG
	70801	CCGGGGAGGA	CGCGGGGAC	TACCA CGGCCG	TGCTGCCGCG	CGGGGAGAGCG
45	70861	CGCGCGT GAC	CGCCCTCGCC	GAGCTGCACG	CCCACGGCGT	CCCGGTCGAC
	70921	TAC1GGCCGG	TGGCCGCCA	GTGGACCTTC	CCGTGTACGC	GTTCCAGCAC
	70981	GGCTGGCCCC	GGCCGTGGCG	GGGGCGCCGG	CCACCGTGGC	GGACACCGGG
	71041	AGTCCGAGCC	GGAGGACCTC	ACCGTCGCGG	AGATCGTCCG	TGCGCGCACC
	71101	TCGGCGTCAC	GGACCCGCC	GACGTCGATG	CGGAAGCGAC	GTCAGCTTCCG
50	71161	ACTCACTGGC	GGTGCAGCGG	CTGCGCAACC	AGCTCGCCTC	GGCAACCGGG
	71221	CGCGGCCGT	CCTGTCGAC	CACGACACCC	CGGCCGCGT	CACCGCGTTC
	71281	GGATCGAGGC	CGGCCAGGAC	CGGATCGAGG	CGGGCGAGGA	CGACGACGCG
	71341	TCTCGCTCCT	GGAGGGAGATG	GAGTCGCTCG	ACGCCGCGGA	CATCGCGCG
	71401	CGGAGCGTGC	GGCCATCGCC	GATCTGCTCG	ACAAGCTCGC	CCATACCTGG
55	71461	GATGAGCACC	GATACGACG	AGGGAAACGCC	GCCCCGCCGG	CGCTGCCCAT
	71521	GGACGGTCAC	CGGCCATCC	TGGAGAGCGG	CACGGTGGGT	TCGTCGAC
	71581	CAAGCACTGG	CTGGTCGCCG	CCGCCGAGGA	CGTCAAGCTG	GTCACCAACG
	71641	CAGCTCGGCC	CGCCCGTCCG	AGATGCTGCC	CGACCCGGCG	CCCGGCTGGT
	71701	GGACTCACCG	GAGCACAA	GCTACCGGGCA	GAAGATCGCG	GGGGACTTCA
60	71761	GGCGCGCAAG	CGGGAGGACT	TCGTCGCCGA	GGCCGCCGAC	GCCTGCCTGG
	71821	GGCCCGGGGA	CCCGGACCG	ACCTCATCCC	CGGGTACGCC	AAGCGCGCTG
	71881	CATCAACGCG	CTGTACGGGC	TCACCCCTGA	GGAGGGGGCC	GTGCTGGAGG
	71941	CGACATCACC	GGCTCGCCG	ATCTGGACAG	CGTCAAGACG	CTGACCGACG
	72001	GCACGCGCTG	CGGCTGGTCC	GGCGAAGCG	TGACGAGCGG	GGCGAGGAGC
	72061	GCTGGCCTCG	GCCGACGACG	GCGAGATCTC	GCTCAGCGAC	GACGAGGCAGA
						CGGGCGTGTT

72121	CGCGACGCTG	CTGTTGCCG	GCCACGACTC	GGTGCAGCAG	ATGGTCGGCT	ACTGCCCTA	
72181	CGCACTGCTC	AGCCACCCCG	AGCAGCAGGC	GGCGCTGCGC	GCGCGCCCCG	AGCTGGTCGA	
72241	CAACCGGGTC	GAGGAGATGC	TCCGTTTCCT	GCCCCGCAAC	CAGATGGGCG	TACCGCGCGT	
5	72301	CTGTGTCGAG	GACGTGATG	TGCGGGCGT	GCGCATCCGT	GCGGGCGACA	ACGTGATCCC
72361	GCTCTACTCG	ACGGCCAACC	GCGACCCCGA	GGTGTCCC	CAGCCCGACA	CCTTCGATGT	
72421	GACGCGCCCG	CTGGAGGGCA	ACCTCGCGTT	CGGCCACGGC	ATTACAAGT	GTCCCGGCCA	
72481	GCACATCGCC	CGGGTGTCA	TCAAGGTCGC	CTGCCTGCGG	TTGTTCGAGC	GTTCGGCGA	
72541	CGTCCGGCTG	GCCGGCGACG	TGCCGATGAA	CGAGGGGCTC	GGGCTGTTCA	GCCCGGCCGA	
10	72601	GCTGCGGGTC	ACCTGGGGG	CGGCATGAGT	CACCCGGTGG	AGACGTTGCG	GTTGCCGAAC
72661	GGGACGACGG	TCGCGCACAT	CAACGCGGGC	GAGGCCAGT	TCCTCTACCG	GGAGATCTTC	
72721	ACCCAGCGCT	GCTACCTGCG	CCACGGTGT	GACCTGCGCC	CGGGGGACGT	GGTGTTCGAC	
72781	GTCGGCGCGA	ACATCGGCAT	GTCACGCTT	TTCGCGCATC	TGGAGTGTCC	TGGTGTGACC	
72841	GTGCACGCCT	TCGAGCCC	GCCC	TTCGCGCGC	TGCGGGCGAA	CGTGACGCGG	
72901	CACGGCATCC	CGGGCCAGGC	GGACCA	GGGGTCTCCG	ACAGCTCCC	CACCCCGAAG	
15	72961	ATGACCTTCT	ATCCCACGC	CACGCTGATG	TCCGGTTCC	ACCGGGATGC	CGCGGCCCGG
73021	ACGGAGCTGT	TGCGCACGCT	CGGCCTCAAC	GGCGGCTACA	CCGCCGAGGA	CGTCGACACC	
73081	ATGCTCGCGC	AACTGCCGA	CGTCAGCGAG	GAGATCGAA	CCCCTGTGGT	CCGGCTCTCC	
73141	GACGTATCG	CGGAGCGCGG	TATCAGG	ATCGGCTGC	TGAAGGTGCA	CGTGGAGAAG	
20	73201	AGCGAACGGC	AGGTCTCGC	CGGCCTCGAG	GACACCGACT	GGCCCCGTAT	CCGCCAGGTC
73261	GTCGGGGAGG	TCCACGACAT	CGACGGCGCG	CTCGAGGAGG	TCGTACGCGT	GTCGGCGGC	
73321	CATGGCTTCA	CCGTGGTCG	CGAGCAGGAA	CCGCTGTTG	CCGGCACG	CATCCACCAG	
73381	GTCGCCGCGC	GGCGGGTGGC	CGGCTGAGCG	CCGTCGGG	CGCGGCCGTC	CGCACCGCG	
73441	GCCGCGGTGC	GGACGGCGC	TCAGCCGGC	TCGGACAGTT	CCTTGGG	TTGCTGACGG	
25	73501	CCCTTCACCC	CCAGCTGCG	GAACACGTTG	GTGAGGTGCT	TTTCCACCGT	GCTGGAGG1G
73561	ACGAACAGCT	GGCTGGCGAT	CTCCTGTTG	GTGCGCCG	CCGCGGCCGTC	CGACGCCACC	
73621	CGCCGCTCCG	CCTCGGT	CGATGTGATC	CGCTGCGCCG	GGCTCACGTC	CTGGGTGCCG	
73681	TCCCGCTCCG	AGGACTCCC	ACCGAGCCG	CGGAGGAGCG	GCACGGCTCC	GCACCTGGTC	
73741	GCGAGGTGCC	GTGCGCCG	GAACAGTCCC	CGCGCACGGC	TGTGCCGCG	GAGCATGCCG	
30	73801	CACGCTTCG	CCATGTCG	GAGGACGCG	GCCAGCTCGT	ACTGGTCGCG	GCACATGATG
73861	AGCAGATCGG	CGGCCTCGC	GAGCAGTTG	ATCCGTTG	CCGGCGGACT	GTAGGCCG	
73921	TGCAACCGCA	GGTCATCAC	CCCGCCCG	GACCCATCG	CCGGGACAG	CTGCTCGGAG	
73981	ATGAGCCTCA	GGCCCTCGC	ACGGCCGCG	CCGAGCAGCA	GAAGCGCTTC	GGCGCGCTCG	
74041	ACCCGCCACA	GGGCCAGGCC	CGGCACGTC	ACGGACCAGC	GTCGATCCG	CTCCCCGAG	
74101	TCCCGGAACG	CGTTGTACG	CGCCCGGTAC	CGCCCGGCG	CGAGATGGT	TTGCCCACGG	
35	74161	GCCAGACCA	TGTGAGTCC	GAAGAGGCTG	TCGGAGGTCT	CCTCCGGCAA	CGGCTCGCG
74221	AGCACCAGCT	CCGCCCCGTC	CAGGTGCGCC	AGTCGGATCG	CGCGGCCAC	GGTGTGCTC	
74281	AGCGGCAATG	CGGCGGCCAT	CCCCCAGGAG	GGCACGACCC	GGGGGGCGAG	CGCGGCCCTCG	
74341	CCGCATTGCA	CGGCGGGCGT	CAGGTGCGCC	CGGCGCAGCG	CGGCCTCGGC	GCGGAACCCC	
40	74401	GCGTGGACCG	CCTCGTCG	CGGGTCCG	ATGTTGTCG	CACCGGCCAG	CTTGTGACCC
74461	CAGGACTGGA	CGGCATCGG	GTCCCTGGCG	TAGAGCAGGG	CCAGCAACGC	CATCATGGTC	
74521	GTGGTCCGGT	CCGTCGTGAC	CCGGGAGTGC	TGGAGCACGT	ACTCGGCTT	GGCCTCGG	
74581	TGTTCGGACC	AGCCGCGCAG	CGCGTTGCTC	AGGGCCTTGT	CGCGACGCG	GCGGTGCCG	
74641	ACGGCTCCGG	AAAACGAGG	GACCTCGTCC	TCGGCCGGCG	GATGGCCCG	ACGCGGCCA	
45	74701	TCGGCCGCGC	CGGGATAGAT	CAGCGCGAGG	GACAGGTCCG	CGACGCGCAG	GTGCGCCCG
74761	CCCTGCTCGC	TCGGGGCGC	GGAGCGCTG	GGCCCGAGGA	CCTCGCGGC	CTCGCCCG	
74821	CGCCCGTCCA	TCGCCAGCCA	GCAGGCGAGC	GACACGGCGT	GTCGCTGGA	GAGGAGCGT	
74881	TCCCGCGACG	CGGTGAGCAG	CTCGGGCACA	TGCCGGCCGG	ATCTGGCG	ATCGCAGAGC	
74941	CGCTCGATGG	CGGCGGTGTC	GACGCGCAGT	CGGGCGTGA	CGCGGGGGTC	GTCGGAGGCC	
50	75001	CGGTAGGCAGA	ACTCCAGGT	GGTGACGGCC	TCGTCGAGCT	CGCCGCGCAG	GTGGTGTCTG
75061	CGCGCCGGCGT	CGGTGAACAG	CCCGGCGAC	TCGGCCCGT	GCACCCGGCC	GGTACCCATC	
75121	TGGTGGCGGG	CGAGCACCTT	GCTGGCCAC	CCGCGGTCCC	CGAGCAGTTC	CAGGCCAGC	
75181	TCTGTCAGGC	CACGCCGCTC	GGCGGGCGAG	AGGTGTCGA	GTACGACGGA	GCGGGCCGCG	
75241	GGGTGCGGG	ACCGCCCTTC	CCGCAGCAGC	CGCCCCTCGA	CCAGCTGTT	GTGGGCTG	
55	75301	TCGACCGCC	CGGTGTCGAG	GCCGGTCATC	CGCTGACGA	GGGTGAGTTC	GACACTCTCG
75361	CCGAGCACCG	CGGAAGCTCG	GGGCAGCCTC	AGCGCGGCCG	GGCGCGAACG	ATAGAGCGAC	
75421	CCGAGGTAGG	CGAGCCGTA	CGCCCCCCCC	GGCACCACTT	CCAGGCCACCC	TGAGGTCCGT	
75481	GTCCGTGCCT	CCCGGATGTC	GTCGATCAGG	CCGTGCG	GGAGCAGGTT	GGCGCCGGTC	
75541	GCCCGGAACG	CCTGGGCCAC	CACGTCGTC	TGCGCGT	GGCCGAGGTT	CCGGCGCACG	
60	75601	AGTCGGTGG	TCTGCGCTC	GGTGAGCGGG	CGCAGCGCGA	TCTCCTGGTA	GTGGCGCAGA
75661	CTCAGCACTG	CCGCCCCGAA	TTGGGAGTGG	GCGGGCGTC	GCCGGAGCAG	CTCGTCAGC	
75721	ACGATGGCGA	CACGGGCCG	GCTGATGCG	CGCGCGAGGT	GGAGCAGGCA	GCGCAGCGAC	
75781	GGCGCGTCGG	CGTGGTGCAC	GTCGTCGATG	CCGATCAGTA	CGGGCCGCTC	CGCGCGCAGC	
75841	GTCAGCACCG	TGCGGGTGG	TTCGGTCCCC	AGCGGTTG	CGACGTCG	CGGCAGGTT	
75901	TCGACGATG	CCGTCAGCCG	GACCAGCTC	GGTGTCCGGG	GGGCCAGCTC	GGGCTGGTC	

75961 AGGAGCTGGC CGAGCATGCC GTACGGCAGG GCCCGCTCCT CCATGGAGCA CACCGCGCGA
 76021 AGGGTGACGA AGCCGGCTT GGCGCGGGG GCGTCGAGGA GTTCGGTCTT GCCGCAGGCG
 76081 ATCGGCCCGG TGACGGCGC GACGACGCC CGCCCGCCCC CCGCTCGGGT GAGCGCCCGG
 76141 TGGAGGGAAC CGAACTCGTC ATCGCGGGCG ATCAGGTCTG GGGGAGATAA GCGCGCTATC
 5 76201 ACGAATGGAA CTACCTCGCG ACCGTCGTGG AAACCCATAG GCATCACATG GCTTGTGAT
 76261 CTGTACGGCT GTGATTCAAG CTGGCGGGAT GCTGTGCTAC AGATGGGAAG ATGTGA^mCTA
 76321 GGGCGTGCC GTTCCCTCA GAGCCGACCG CCCCCGGCGC CACCCGCCGT ACCCCCTGGG
 76381 CCACCAGCTC GGCGACCCGC TCCTGGTGGT CGACGAGGTA GAAGTGCCCG CGGGGAAGA
 76441 CCTCCACCGT GGTCGGCGCG GTCTGTGCCC CGGCCCCAGGC GTGGGCCTGC TCCACCGTCG
 10 76501 TCTTCGGATC GTCGTACCG ATGCACACCG TGATCGGCGT CTCCAGCGGC GGCGCGGGCT
 76561 CCCACCGGTA CGTCTCCGCC GCGTAGTAGT CCGCCCGCAA CGGCGCCAGG ATCAGCGCGC
 76621 GCATTTGTC GTCGCCATC ACATCGGCGC TCGTCCCGCC GAGGCCGATG ACCGCCGCCA
 76681 GCAGCTCGTC GTCGGACCGG AGGTGGTCTT GGTCGGCGCG CGGCTGCGAC GGCGCCCGCC
 15 76741 GGCCCGAGAC GATCAGGTGC GCCACCGGGA GCCGCTGGGC CAGCTGAAC GCGAGTGTGCG
 76801 CGCCCATGCT GTGGCGAAC AGCACCAGCG GACGGTCCAG CCCCAGGCTTC AACGCCCTCGG
 76861 CCACGAGGCC GGCGAGAACA CGCAGGTGCG GCACCGCCTC CTCGTCGCGG CGGTCTGGC
 76921 GGCGGGGTA CTGCACGGCG TACACGTCCG CCACCGGGGC GAGCGCACCG GCCAGCGAA
 76981 GGTAGAACGT CGCCGATCCG CCGGCGTGGG GCAGCAGCAC CACCCGTACC GGGGCCTCGG
 77041 GCGTGGGGAA GAACTGCCGC AGCCAGAGTT CCGAGCTCAC CGCACCCCCCT CGGCCGCGAC
 20 77101 CTGGGGAGCC CGGAACCGGG TGATCTCGGC CAAGTGTCTC TCCCGCATCT CGGGTCGCGT
 77161 CACGCCCAT CCCTCCCTCG GCGCCAGACAA GAGGACGCCG ACTTTGCCGT TGTGACATT
 77221 GCGATGCACA TCGCGGACCG CGGACCCGAC GTCTCGAGC GGGTAGGTCA CGCACAGCGT
 77281 CGGGTGCACC ATCCCCCTTG AGATCAGGCG GTTCGCTC CACGCCTCAC GATACTTCGC
 77341 GAAAGTGGGTA CCGATGATCC GCTTCACCGGA CATCCACAGG TACCGATTGT CAAAGGCGTG
 25 77401 CTCGTATCCC GAGGTTGACG CGCAGGTGAC GATCGTGCCA CCCCCGACGTG TCACGT^mGAC
 77461 ACTCGCGCCG AACGTCCGCC GCCCGGGGTG CTCGAACACCG ATGTCGGGAT CGTCACCGCC
 77521 GGTCAAGCTCC CGGATC

Those of skill in the art will recognize that, due to the degenerate nature of the
 30 genetic code, a variety of DNA compounds differing in their nucleotide sequences can be
 used to encode a given amino acid sequence of the invention. The native DNA sequence
 encoding the FK-520 PKS of *Streptomyces hygroscopicus* is shown herein merely to
 illustrate a preferred embodiment of the invention, and the present invention includes
 DNA compounds of any sequence that encode the amino acid sequences of the
 35 polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically
 tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid
 sequence without loss or significant loss of a desired activity. The present invention
 includes such polypeptides with alternate amino acid sequences, and the amino acid
 sequences shown merely illustrate preferred embodiments of the invention.

40 The recombinant nucleic acids, proteins, and peptides of the invention are many
 and diverse. To facilitate an understanding of the invention and the diverse compounds
 and methods provided thereby, the following general description of the FK-520 PKS
 genes and modules of the PKS proteins encoded thereby is provided. This general
 description is followed by a more detailed description of the various domains and
 45 modules of the FK-520 PKS contained in and encoded by the compounds of the
 invention. In this description, reference to a heterologous PKS refers to any PKS other
 than the FK-520 PKS. Unless otherwise indicated, reference to a PKS includes reference

to a portion of a PKS. Moreover, reference to a domain, module, or PKS includes reference to the nucleic acids encoding the same and vice-versa, because the methods and reagents of the invention provide or enable one to prepare proteins and the nucleic acids that encode them.

5 The FK-520 PKS is composed of three proteins encoded by three genes designated *fkbA*, *fkbB*, and *fkbC*. The *fkbA* ORF encodes extender modules 7 - 10 of the PKS. The *fkbB* ORF encodes the loading module (the CoA ligase) and extender modules 1 - 4 of the PKS. The *fkbC* ORF encodes extender modules 5 - 6 of the PKS. The *fkbP* ORF encodes the NRPS that attaches the pipecolic acid and cyclizes the FK-520 10 polyketide.

15 The loading module of the FK-520 PKS includes a CoA ligase, an ER domain, and an ACP domain. The starter building block or unit for FK-520 is believed to be a dihydroxycyclohexene carboxylic acid, which is derived from shikimate. The recombinant DNA compounds of the invention that encode the loading module of the 20 FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of methods and in a variety of compounds. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for the loading module of the heterologous PKS is replaced by the coding sequence for the FK-520 loading module, provides a novel PKS coding sequence. Examples of heterologous PKS coding 25 sequences include the rapamycin, FK-506, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

30 In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, either replacing the CoA ligase with a different CoA ligase, deleting the ER, or replacing the ER with a different ER. In addition, or alternatively, the ACP can be replaced by another ACP. In similar fashion, the corresponding domains in another loading or extender module can be replaced by one or more domains of the FK-520 PKS. The resulting heterologous loading module coding sequence can be utilized in conjunction with a coding sequence for a PKS that 35 synthesizes FK-520, an FK-520 derivative, or another polyketide.

The first extender module of the FK-520 PKS includes a KS domain, an AT domain specific for methylmalonyl CoA, a DH domain, a KR domain, and an ACP domain. The recombinant DNA compounds of the invention that encode the first extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the first extender module of the FK-520 PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the FK-520 PKS is inserted into a DNA compound that comprises the remainder of the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or only a portion of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting either the DH or KR or both; replacing the DH or KR or both with another DH or KR; and/or inserting an ER. In replacing or inserting KR, DH, and ER domains, it is often beneficial to replace the existing KR, DH, and ER domains with the complete set of domains desired from another module. Thus, if one desires to insert an ER domain, one may simply replace the existing KR and DH domains with a KR, DH, and ER set of domains from a module containing such domains. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a gene for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous first extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the first extender module of the FK-520 PKS.

In an illustrative embodiment of this aspect of the invention, the invention provides recombinant PKSs and recombinant DNA compounds and vectors that encode

such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are especially useful when placed in translational reading frame with the remaining modules and domains of an FK-520 or FK-520 derivative PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the PKS 5 encoded thereby can be fed or supplied with N-acylcysteamine thioesters of novel precursor molecules to prepare FK-520 derivatives. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT patent publication Nos. US97/02358 and US99/03986, each of which is incorporated herein by reference.

The second extender module of the FK-520 PKS includes a KS, an AT specific 10 for methylmalonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the second extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 second extender module is inserted into a DNA compound that comprises the 15 coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the second 20 extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the second extender module coding 25 sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of 30 these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous second extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK- 35 520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding

domains in a module of a heterologous PKS can be replaced by one or more domains of the second extender module of the FK-520 PKS.

The third extender module of the FK-520 PKS includes a KS, an AT specific for malonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of 5 the invention that encode the third extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence 10 for a module of the heterologous PKS is either replaced by that for the third extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding 15 sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 20 malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of 25 these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK-30 520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the third extender module of the FK-520 PKS.

The fourth extender module of the FK-520 PKS includes a KS, an AT that binds ethylmalonyl CoA, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the fourth extender module of the FK-520 PKS and the 35 corresponding polypeptides encoded thereby are useful for a variety of applications. In

one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fourth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender 5 module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the fourth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the remainder of the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS 10 that produces an FK-520 derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the ethylmalonyl CoA specific AT with a malonyl CoA, methylmalonyl CoA, or 2-hydroxymalonyl CoA 15 specific AT; and/or deleting the inactive DH, inserting a KR, a KR and an active DH, or a KR, an active DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, a PKS for a polyketide other than FK-520, or from 20 chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fourth extender module of the FK-520 PKS.

25 As illustrative examples, the present invention provides recombinant genes, vectors, and host cells that result from the conversion of the FK-506 PKS to an FK-520 PKS and vice-versa. In one embodiment, the invention provides a recombinant set of FK-506 PKS genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by 30 those for the AT domain of the fourth extender module of the FK-520 PKS. This recombinant PKS can be used to produce FK-520 in recombinant host cells. In another embodiment, the invention provides a recombinant set of FK-520 PKS genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by those for the AT domain of

the fourth extender module of the FK-506 PKS. This recombinant PKS can be used to produce FK-506 in recombinant host cells.

Other examples of hybrid PKS enzymes of the invention include those in which the AT domain of module 4 has been replaced with a malonyl specific AT domain to 5 provide a PKS that produces 21-desethyl-FK520 or with a methylmalonyl specific AT domain to provide a PKS that produces 21-desethyl-21-methyl-FK520. Another hybrid PKS of the invention is prepared by replacing the AT and inactive KR domain of FK-520 extender module 4 with a methylmalonyl specific AT and an active KR domain, such as, for example, from module 2 of the DEBS or oleandolide PKS enzymes, to produce 21-10 desethyl-21-methyl-22-desoxo-22-hydroxy-FK520. The compounds produced by these hybrid PKS enzymes are neurotrophins.

The fifth extender module of the FK-520 PKS includes a KS, an AT that binds methylmalonyl CoA, a DH, a KR, and an ACP. The recombinant DNA compounds of 15 the invention that encode the fifth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fifth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fifth extender 20 module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fifth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

25 In another embodiment, a portion of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one or both of the DH and KR; replacing 30 any one or both of the DH and KR with either a KR and/or DH; and/or inserting an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from 35 chemical synthesis. The resulting heterologous fifth extender module coding sequence

can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fifth extender module of the FK-520 PKS.

5 In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the DH domain of the fifth extender module have been deleted or mutated to render the DH non-functional. In one such mutated gene, the KR and DH coding sequences are replaced with those encoding only a KR domain from another PKS gene. The resulting PKS genes code for the
10 expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-19 to C-20 double bond of FK-520 and has a C-20 hydroxyl group. Such analogs are preferred neurotrophins, because they have little or no immunosuppressant activity. This recombinant fifth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment,
15 the present invention provides a recombinant FK-520 PKS that contains both this fifth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that
20 express this recombinant PKS and so synthesize the corresponding (lacking the C-19 to C-20 double bond of FK-506 and having a C-20 hydroxyl group) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the DH domain of module 5 has been deleted or otherwise rendered inactive and thus produces this novel polyketide.

25 The sixth extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the sixth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes
30 the FK-520 sixth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In
35 another embodiment, a DNA compound comprising a sequence that encodes the sixth

extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the sixth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one, two, or all three of the KR, DH, and ER; and/or replacing any one, two, or all three of the KR, DH, and ER with another KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous sixth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the sixth extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the DH and ER domains of the sixth extender module have been deleted or mutated to render them non-functional. In one such mutated gene, the KR, ER, and DH coding sequences are replaced with those encoding only a KR domain from another PKS gene. This can also be accomplished by simply replacing the coding sequences for extender module six with those for an extender module having a methylmalonyl specific AT and only a KR domain from a heterologous PKS gene, such as, for example, the coding sequences for extender module two encoded by the *eryA1* gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that has a C-18 hydroxyl group. Such analogs are preferred neurotrophins, because they have little or no immunosuppressant activity. This recombinant sixth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this sixth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-

506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (having a C-18 hydroxyl group) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the DH and ER domains of module 6 have 5 been deleted or otherwise rendered inactive and thus produces this novel polyketide.

The seventh extender module of the FK-520 PKS includes a KS, an AT specific for 2-hydroxymalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the seventh extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of 10 applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 seventh extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the seventh extender module of the FK-520 PKS or the latter is merely added to coding 15 sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the seventh extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

20 In another embodiment, a portion or all of the seventh extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 2-hydroxymalonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or malonyl CoA specific AT; deleting the KR, the DH, and/or the ER; and/or replacing the 25 KR, DH, and/or ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous 30 seventh extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the seventh extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the AT domain of the seventh extender module has been replaced with those encoding an AT domain for malonyl, methylmalonyl, or ethylmalonyl CoA from another PKS gene. The resulting PKS genes 5 code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-15 methoxy group, having instead a hydrogen, methyl, or ethyl group at that position, respectively. Such analogs are preferred, because they are more slowly metabolized than FK-520. This recombinant seventh extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an 10 illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this seventh extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK- 15 506 but that express this recombinant PKS and so synthesize the corresponding (C-15-desmethoxy) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the AT domain of module 7 has been replaced and thus produces this novel polyketide.

In another illustrative embodiment, the present invention provides a hybrid PKS 20 in which the AT and KR domains of module 7 of the FK-520 PKS are replaced by a methylmalonyl specific AT domain and an inactive KR domain, such as, for example, the AT and KR domains of extender module 6 of the rapamycin PKS. The resulting hybrid PKS produces 15-desmethoxy-15-methyl-16-oxo-FK-520, a neurotrophin compound.

25 The eighth extender module of the FK-520 PKS includes a KS, an AT specific for 2-hydroxymalonyl CoA, a KR, and an ACP. The recombinant DNA compounds of the invention that encode the eighth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 30 eighth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the eighth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another 35 embodiment, a DNA compound comprising a sequence that encodes the eighth extender

module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

- In another embodiment, a portion of the eighth extender module coding sequence
- 5 is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 2-hydroxymalonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or malonyl CoA specific AT; deleting or replacing the KR; and/or inserting a DH or a DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP.
- 10 In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous eighth extender module coding sequence can be utilized in conjunction with a PKS that synthesizes FK-520, an FK-520
- 15 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the eighth extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the AT domain of the eighth extender module has been replaced with those encoding an AT domain for malonyl, methylmalonyl, or ethylmalonyl CoA from another PKS gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-13 methoxy group, having instead a hydrogen, methyl, or ethyl group at that position, respectively. Such analogs are preferred, because they are more slowly metabolized than FK-520. This recombinant eighth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this eighth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (C-13-desmethoxy) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the AT domain of module 8 has been replaced and thus produces this novel polyketide.

- The ninth extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the ninth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of
- 5 applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 ninth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the ninth extender module of the FK-520 PKS or the latter is merely added to coding sequences
- 10 for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the ninth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.
- 15 In another embodiment, a portion of the ninth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one, two, or all three of the KR, DH, and
- 20 ER; and/or replacing any one, two, or all three of the KR, DH, and ER with another KR, DH, and/or ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or
- 25 from chemical synthesis. The resulting heterologous ninth extender module coding sequence can be utilized in conjunction with a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the ninth extender module of the FK-520 PKS.
- 30 The tenth extender module of the FK-520 PKS includes a KS, an AT specific for malonyl CoA, and an ACP. The recombinant DNA compounds of the invention that encode the tenth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 tenth extender module is inserted into a DNA compound that comprises the coding sequence

for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the tenth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a 5 DNA compound comprising a sequence that encodes the tenth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion or all of the tenth extender module coding 10 sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; and/or inserting a KR, a KR and DH, or a KR, DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. 15 In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous tenth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that 20 synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the tenth extender module of the FK-520 PKS.

The FK-520 polyketide precursor produced by the action of the tenth extender 25 module of the PKS is then attached to pipecolic acid and cyclized to form FK-520. The enzyme FkbP is the NRPS like enzyme that catalyzes these reactions. FkbP also includes a thioesterase activity that cleaves the nascent FK-520 polyketide from the NRPS. The present invention provides recombinant DNA compounds that encode the *fkbP* gene and so provides recombinant methods for expressing the *fkbP* gene product in recombinant host cells. The recombinant *fkbP* genes of the invention include those in which the 30 coding sequence for the adenylation domain has been mutated or replaced with coding sequences from other NRPS like enzymes so that the resulting recombinant FkbP incorporates a moiety other than pipecolic acid. For the construction of host cells that do not naturally produce pipecolic acid, the present invention provides recombinant DNA compounds that express the enzymes that catalyze at least some of the biosynthesis of 35 pipecolic acid (see Nielsen *et al.*, 1991, *Biochem.* 30: 5789-96). The *fkbL* gene encodes a

homolog of RapL, a lysine cyclodeaminase responsible in part for producing the pipecolate unit added to the end of the polyketide chain. The *fkB* and *fkL* recombinant genes of the invention can be used in heterologous hosts to produce compounds such as FK-520 or, in conjunction with other PKS or NRPS genes, to produce known or novel 5 polyketides and non-ribosomal peptides.

The present invention also provides recombinant DNA compounds that encode the P450 oxidase and methyltransferase genes involved in the biosynthesis of FK-520. Figure 2 shows the various sites on the FK-520 polyketide core structure at which these enzymes act. By providing these genes in recombinant form, the present invention 10 provides recombinant host cells that can produce FK-520. This is accomplished by introducing the recombinant PKS, P450 oxidase, and methyltransferase genes into a heterologous host cell. In a preferred embodiment, the heterologous host cell is *Streptomyces coelicolor* CH999 or *Streptomyces lividans* K4-114, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 15 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference. In addition, by providing recombinant host cells that express only a subset of these genes, the present invention provides methods for making FK-520 precursor compounds not readily obtainable by other means.

In a related aspect, the present invention provides recombinant DNA compounds 20 and vectors that are useful in generating, by homologous recombination, recombinant host cells that produce FK-520 precursor compounds. In this aspect of the invention, a native host cell that produces FK-520 is transformed with a vector (such as an SCP2* derived vector for *Streptomyces* host cells) that encodes one or more disrupted genes (i.e., a hydroxylase, a methyltransferase, or both) or merely flanking regions from those 25 genes. When the vector integrates by homologous recombination, the native, functional gene is deleted or replaced by the non-functional recombinant gene, and the resulting host cell thus produces an FK-520 precursor. Such host cells can also be complemented by introduction of a modified form of the deleted or mutated non-functional gene to produce a novel compound.

30 In one important embodiment, the present invention provides a hybrid PKS and the corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the present invention a hybrid PKS is a recombinant PKS that comprises all or part of one or more modules and thioesterase/cyclase domain of a first PKS and all or part of one or more modules, loading module, and thioesterase/cyclase

domain of a second PKS. In one preferred embodiment, the first PKS is all or part of the FK-520 PKS, and the second PKS is only a portion or all of a non-FK-520 PKS.

One example of the preferred embodiment is an FK-520 PKS in which the AT domain of module 8, which specifies a hydroxymalonyl CoA and from which the C-13 5 methoxy group of FK-520 is derived, is replaced by an AT domain that specifies a malonyl, methylmalonyl, or ethylmalonyl CoA. Examples of such replacement AT domains include the AT domains from modules 3, 12, and 13 of the rapamycin PKS and from modules 1 and 2 of the erythromycin PKS. Such replacements, conducted at the level of the gene for the PKS, are illustrated in the examples below. Another 10 illustrative example of such a hybrid PKS includes an FK-520 PKS in which the natural loading module has been replaced with a loading module of another PKS. Another example of such a hybrid PKS is an FK-520 PKS in which the AT domain of module three is replaced with an AT domain that binds methylmalonyl CoA.

In another preferred embodiment, the first PKS is most but not all of a non-FK- 15 520 PKS, and the second PKS is only a portion or all of the FK-520 PKS. An illustrative example of such a hybrid PKS includes an erythromycin PKS in which an AT specific for methylmalonyl CoA is replaced with an AT from the FK-520 PKS specific for malonyl CoA.

Those of skill in the art will recognize that all or part of either the first or second 20 PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See U.S. provisional patent application Serial No. 60/091,526, incorporated herein by reference. The state of the art in DNA synthesis allows the artisan to construct *de novo* 25 DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. For purposes of the present invention, such synthetic DNA compounds are deemed to be a portion of a PKS.

Thus, the hybrid modules of the invention are incorporated into a PKS to provide a hybrid PKS of the invention. A hybrid PKS of the invention can result not only:

(i) from fusions of heterologous domain (where heterologous means the domains 30 in that module are from at least two different naturally occurring modules) coding sequences to produce a hybrid module coding sequence contained in a PKS gene whose product is incorporated into a PKS,
but also:

(ii) from fusions of heterologous module (where heterologous module means two 35 modules are adjacent to one another that are not adjacent to one another in naturally

occurring PKS enzymes) coding sequences to produce a hybrid coding sequence contained in a PKS gene whose product is incorporated into a PKS,

- (iii) from expression of one or more FK-520 PKS genes with one or more non-FK-520 PKS genes, including both naturally occurring and recombinant non-FK-520
5 PKS genes, and
(iv) from combinations of the foregoing.

Various hybrid PKSs of the invention illustrating these various alternatives are described herein.

Examples of the production of a hybrid PKS by co-expression of PKS genes from
10 the FK-520 PKS and another non-FK-520 PKS include hybrid PKS enzymes produced by coexpression of FK-520 and rapamycin PKS genes. Preferably, such hybrid PKS enzymes are produced in recombinant *Streptomyces* host cells that produce FK-520 or FK-506 but have been mutated to inactivate the gene whose function is to be replaced by the rapamycin PKS gene introduced to produce the hybrid PKS. Particular examples
15 include (i) replacement of the *fkbC* gene with the *rapB* gene; and (ii) replacement of the *fkbA* gene with the *rapC* gene. The latter hybrid PKS produces 13,15-didesmethoxy-FK-520, if the host cell is an FK-520 producing host cell, and 13,15-didesmethoxy-FK-506, if the host cell is an FK-506 producing host cell. The compounds produced by these hybrid PKS enzymes are immunosuppressants and neurotrophins but can be readily
20 modified to act only as neurotrophins, as described in Example 6, below.

Other illustrative hybrid PKS enzymes of the invention are prepared by replacing the *fkbA* gene of an FK-520 or FK-506 producing host cell with a hybrid *fkbA* gene in which: (a) the extender module 8 through 10, inclusive, coding sequences have been replaced by the coding sequences for extender modules 12 to 14, inclusive, of the
25 rapamycin PKS; and (b) the module 8 coding sequences have been replaced by the module 8 coding sequence of the rifamycin PKS. When expressed with the other, naturally occurring FK-520 or FK-506 PKS genes and the genes of the modification enzymes, the resulting hybrid PKS enzymes produce, respectively, (a) 13-desmethoxy-FK-520 or 13-desmethoxy-FK-506; and (b) 13-desmethoxy-13-methyl-FK-520 or 13-
30 desmethoxy-13-methyl-FK-506. In a preferred embodiment, these recombinant PKS genes of the invention are introduced into the producing host cell by a vector such as pHU204, which is a plasmid pRM5 derivative that has the well-characterized SCP2* replicon, the *colE1* replicon, the *tsr* and *bla* resistance genes, and a *cos* site. This vector can be used to introduce the recombinant *fkbA* replacement gene in an FK-520 or FK-
35 506 producing host cell (or a host cell derived therefrom in which the endogenous *fkbA*

gene has either been rendered inactive by mutation, deletion or homologous recombination with the gene that replaces it) to produce the desired hybrid PKS.

In constructing hybrid PKSs of the invention, certain general methods may be helpful. For example, it is often beneficial to retain the framework of the module to be altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a module, it is often preferred to replace the KR domain of the original module with a KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of modular polyketide synthases in the choice and stereochemical fate of extender units," *Biochemistry* 38(5):1643-1651, incorporated herein by reference. Stereochemistry can also be changed by changing the KR domain. Also, one can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can also take advantage of known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases," *Science* 284: 482-485, incorporated herein by reference.

The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the present invention.*

Avermectin

25 U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemalectin.

30 MacNeil *et al.*, 1992, *Gene* 115: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

Ikeda *et al.*, Aug. 1999, Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*, *Proc. Natl. Acad. Sci. USA* 96: 9509-9514.

35 **Candididin (FR008)**

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

Epothilone

U.S. Pat. App. Serial No. 60/130,560, filed 22 April 1999.

Erythromycin

5 PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science* 252:675-9.

10 Cortes *et al.*, 8 Nov. 1990, *Nature* 348:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of *Saccharopolyspora erythraea*.

Glycosylation Enzymes

PCT Pat. App. Pub. No. 97/23630 to Abbott.

FK-506

15 Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK-506, *Eur. J. biochem.* 256: 528-534.

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK-506, *Eur. J. Biochem.* 244: 74-80.

Methyltransferase

20 US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858. 31-O-desmethyl-FK-506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK-506 and FK-520, *J. Bacteriol.* 178: 5243-5248.

25 *Streptomyces hygroscopicus*

U.S. patent application Serial No. 09/154,083, filed 16 Sep. 1998.

Lovastatin

U.S. Pat. No. 5,744,350 to Merck.

Narbomycin

30 U.S. patent application Serial No. 60/107,093, filed 5 Nov. 1998, and Serial No. 60/120,254, filed 16 Feb. 1999.

Nemadectin

MacNeil *et al.*, 1993, *supra*.

Niddamycin

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, *J. Bacteriol.* 179: 7515-7522.

Oleandomycin

- Swan *et al.*, 1994, Characterisation of a *Streptomyces antibioticus* gene encoding 5 a type I polyketide synthase which has an unusual coding sequence, *Mol. Gen. Genet.* 242: 358-362.

U.S. patent application Serial No. 60/120,254, filed 16 Feb. 1999.

- Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases 10 responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet.* 259(3): 299-308.

Picromycin

PCT patent application US99/15047, filed 2 Jul. 1999.

- Xue *et al.*, 1998, Hydroxylation of macrolactones YC-17 and narbomycin is 15 mediated by the *pikC*-encoded cytochrome P450 in *Streptomyces venezuelae*, *Chemistry & Biology* 5(11): 661-667.

Xue *et al.*, Oct. 1998, A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity, *Proc. Natl. Acad. Sci. USA* 95: 12111 12116.

20 **Platenolide**

EP Pat. App. Pub. No. 791,656 to Lilly.

Rapamycin

Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA* 92:7839-7843.

- 25 Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene* 169: 9-16.

Rifamycin

- August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: 30 deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S669, *Chemistry & Biology*, 5(2): 69-79.

Sorangium PKS

U.S. patent application Serial No. 09/144,085, filed 31 Aug. 1998.

Soraphen

- 35 U.S. Pat. No. 5,716,849 to Novartis.

Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

5 **Spiramycin**

U.S. Pat. No. 5,098,837 to Lilly.

Activator Gene

U.S. Pat. No. 5,514,544 to Lilly.

Tylosin

10 EP Pub. No. 791,655 to Lilly.

U.S. Pat. No. 5,876,991 to Lilly.

Kuhstoss *et al.*, 1996, *Gene* 183:231-6., Production of a novel polyketide through the construction of a hybrid polyketide synthase.

Tailoring enzymes

15 Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355. Analysis of five tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there are a wide variety of polyketide synthase genes that serve as readily available sources of DNA and sequence information for use in 20 constructing the hybrid PKS-encoding DNA compounds of the invention. Methods for constructing hybrid PKS-encoding DNA compounds are described without reference to the FK-520 PKS in PCT patent publication No. 98/51695; U.S. Patent Nos. 5,672,491 and 5,712,146 and U.S. patent application Serial Nos. 09/073,538, filed 6 May 1998, and 09/141,908, filed 28 Aug 1998, each of which is incorporated herein by reference.

25 The hybrid PKS-encoding DNA compounds of the invention can be and often are hybrids of more than two PKS genes. Moreover, there are often two or more modules in the hybrid PKS in which all or part of the module is derived from a second (or third) PKS. Thus, as one illustrative example, the present invention provides a hybrid FK-520 PKS that contains the naturally occurring loading module and FkbP as well as modules 30 one, two, four, six, seven, and eight, nine, and ten of the FK-520 PKS and further contains hybrid or heterologous modules three and five. Hybrid or heterologous module three contains an AT domain that is specific of methylmalonyl CoA and can be derived for example, from the erythromycin or rapamycin PKS genes. Hybrid or heterologous module five contains an AT domain that is specific for malonyl CoA and can be derived 35 for example, from the picromycin or rapamycin PKS genes.

While an important embodiment of the present invention relates to hybrid PKS enzymes and corresponding genes, the present invention also provides recombinant FK-520 PKS genes in which there is no second PKS gene sequence present but which differ from the FK-520 PKS gene by one or more deletions. The deletions can encompass one or more modules and/or can be limited to a partial deletion within one or more modules. When a deletion encompasses an entire module, the resulting FK-520 derivative is at least two carbons shorter than the gene from which it was derived. When a deletion is within a module, the deletion typically encompasses a KR, DH, or ER domain, or both DH and ER domains, or both KR and DH domains, or all three KR, DH, and ER domains.

To construct a hybrid PKS or FK-520 derivative PKS gene of the invention, one can employ a technique, described in PCT Pub. No. 98/27203 and U.S. patent application Serial No. 08/989,332, filed 11 Dec. 1997, each of which is incorporated herein by reference, in which the large PKS gene is divided into two or more, typically three, segments, and each segment is placed on a separate expression vector. In this manner, each of the segments of the gene can be altered, and various altered segments can be combined in a single host cell to provide a recombinant PKS gene of the invention. This technique makes more efficient the construction of large libraries of recombinant PKS genes, vectors for expressing those genes, and host cells comprising those vectors.

Thus, in one important embodiment, the recombinant DNA compounds of the invention are expression vectors. As used herein, the term expression vector refers to any nucleic acid that can be introduced into a host cell or cell-free transcription and translation medium. An expression vector can be maintained stably or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a replicating vector in the cytoplasm. An expression vector also comprises a gene that serves to produce RNA that is translated into a polypeptide in the cell or cell extract. Furthermore, expression vectors typically contain additional functional elements, such as resistance-conferring genes to act as selectable markers.

The various components of an expression vector can vary widely, depending on the intended use of the vector. In particular, the components depend on the host cell(s) in which the vector will be used or is intended to function. Vector components for expression and maintenance of vectors in *E. coli* are widely known and commercially available, as are vector components for other commonly used organisms, such as yeast cells and *Streptomyces* cells.

In a preferred embodiment, the expression vectors of the invention are used to construct recombinant *Streptomyces* host cells that express a recombinant PKS of the invention. Preferred *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999 and *S. lividans* K4-114 host cells, which do not produce 5 actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference.

The present invention provides a wide variety of expression vectors for use in 10 *Streptomyces*. For replicating vectors, the origin of replication can be, for example and without limitation, a low copy number vector, such as SCP2* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985); Lydiate *et al.*, 1985, *Gene* 35: 223-235; and Kieser and Melton, 1988, *Gene* 65: 83-91, each of which is incorporated herein by reference), 15 SLP1.2 (Thompson *et al.*, 1982, *Gene* 20: 51-62, incorporated herein by reference), and SG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet.* 219: 341-348, and Bierman *et al.*, 1992, *Gene* 116: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J. Gen. Microbiol.* 129: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol.* 171: 5782-5781; and Servin-Gonzalez, 1993, 20 *Plasmid* 30: 131-140, each of which is incorporated herein by reference). Generally, however, high copy number vectors are not preferred for expression of genes contained on large segments of DNA. For non-replicating and integrating vectors, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p1I, and pBR. 25 For phage based vectors, the phages phiC31 and KC515 can be employed (see Hopwood *et al.*, *supra*).

Typically, the expression vector will comprise one or more marker genes by 30 which host cells containing the vector can be identified and/or selected. Useful antibiotic resistance conferring genes for use in *Streptomyces* host cells include the *ermE* (confers resistance to erythromycin and other macrolides and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA* (confers resistance to spectinomycin and streptomycin), *aacC4* (confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes.

The recombinant PKS gene on the vector will be under the control of a promoter, 35 typically with an attendant ribosome binding site sequence. The present invention

provides the endogenous promoters of the FK-520 PKS and related biosynthetic genes in recombinant form, and these promoters are preferred for use in the native hosts and in heterologous hosts in which the promoters function. A preferred promoter of the invention is the *fkbO* gene promoter, comprised in a sequence of about 270 bp between 5 the start of the open reading frames of the *fkbO* and *fkbB* genes. The *fkbO* promoter is believed to be bi-directional in that it promotes transcription of the genes *fkbO*, *fkbP*, and *fkbA* in one direction and *fkbB*, *fkbC*, and *fkbL* in the other. Thus, in one aspect, the present invention provides a recombinant expression vector comprising the promoter of the *fkbO* gene of an FK-520 producing organism positioned to transcribe a gene other 10 than *fkbO*. In a preferred embodiment the transcribed gene is an FK-520 PKS gene. In another preferred embodiment, the transcribed gene is a gene that encodes a protein comprised in a hybrid PKS.

Heterologous promoters can also be employed and are preferred for use in host 15 cells in which the endogenous FK-520 PKS gene promoters do not function or function poorly. A preferred heterologous promoter is the *actII* promoter and its attendant activator gene *actII-ORF4*, which is provided in the pRM1 and pRM5 expression vectors, *supra*. This promoter is activated in the stationary phase of growth when secondary metabolites are normally synthesized. Other useful *Streptomyces* promoters include without 20 limitation those from the *ermE* gene and the *melC1* gene, which act constitutively, and the *tipA* gene and the *merA* gene, which can be induced at any growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA* promoter, and the gene of interest is 25 placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to enhance the activity of a promoter. Activator genes in addition to the *actII-ORF4* gene discussed above include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, *supra*) to activate promoters under their control.

30 In addition to providing recombinant DNA compounds that encode the FK-520 PKS, the present invention also provides DNA compounds that encode the ethylmalonyl CoA and 2-hydroxymalonyl CoA utilized in the synthesis of FK-520. Thus, the present invention also provides recombinant host cells that express the genes required for the biosynthesis of ethylmalonyl CoA and 2-hydroxymalonyl CoA. Figures 3 and 4 show the

location of these genes on the cosmids of the invention and the biosynthetic pathway that produces ethylmalonyl CoA.

For 2-hydroxymalonyl CoA biosynthesis, the *fkbH*, *fkbI*, *fkbJ*, and *fkbK* genes are sufficient to confer this ability on *Streptomyces* host cells. For conversion of 2-hydroxymalonyl to 2-methoxymalonyl, the *fkbG* gene is also employed. While the complete coding sequence for *fkbH* is provided on the cosmids of the invention, the sequence for this gene provided herein may be missing a T residue, based on a comparison made with a similar gene cloned from the ansamitocin gene cluster by Dr. H. Floss. Where the sequence herein shows one T, there may be two, resulting in an extension of the *fkbH* reading frame to encode the amino acid sequence:

5 MTIVKCLVWDLNTLWRGTVLEDDEVVLTDEIREVITLDDRGILQAVASKNDH
DLAWERLERLGVAEYFVLARIGWGPKSQSVREIATELNFAPTTIAFIDDQPAERA
10 EVAFHLPEVRCYPAEQAATLLSLPEFSPPVSTVDSRRRLMYQAGFARDQAREA
YSGPDEDFLRSLLSMTIAPAGEEELSRRVEELRLTSQMNATGVHYSADLRAL
15 LTDPAHEVLVVTMGDRFGPHGAVGIILLEKKPSTWHLKLLATSCRVVSFGAGAT
ILNWLTDQGARAGAHLVADFRRTDRNRMMEIAYRFAGFADSDCPCVSEVAGAS
AAGVERLHLEPSARPAPTTLTAADIAPVTVSAAG.

For ethylmalonyl CoA biosynthesis, one requires only a crotonyl CoA reductase, which can be supplied by the host cell but can also be supplied by recombinant expression of the *fkbS* gene of the present invention. To increase yield of ethylmalonyl CoA, one can also express the *fkbE* and *fkbU* genes as well. While such production can be achieved using only the recombinant genes above, one can also achieve such production by placing into the recombinant host cell a large segment of the DNA provided by the cosmids of the invention. Thus, for 2-hydroxymalonyl and 2-methoxymalonyl CoA biosynthesis, one can simply provide the cells with the segment of DNA located on the left side of the FK-520 PKS genes shown in Figure 1. For ethylmalonyl CoA biosynthesis, one can simply provide the cells with the segment of DNA located on the right side of the FK-520 PKS genes shown in Figure 1 or, alternatively, both the right and left segments of DNA.

30 The recombinant DNA expression vectors that encode these genes can be used to construct recombinant host cells that can make these important polyketide building blocks from cells that otherwise are unable to produce them. For example, *Streptomyces coelicolor* and *Streptomyces lividans* do not synthesize ethylmalonyl CoA or 2-hydroxymalonyl CoA. The invention provides methods and vectors for constructing recombinant *Streptomyces coelicolor* and *Streptomyces lividans* that are able to

synthesize either or both ethylmalonyl CoA and 2-hydroxymalonyl CoA. These host cells are thus able to make polyketides, those requiring these substrates, that cannot otherwise be made in such cells.

In a preferred embodiment, the present invention provides recombinant

- 5 *Streptomyces* host cells, such as *S. coelicolor* and *S. lividans*, that have been transformed with a recombinant vector of the invention that codes for the expression of the ethylmalonyl CoA biosynthetic genes. The resulting host cells produce ethylmalonyl CoA and so are preferred host cells for the production of polyketides produced by PKS enzymes that comprise one or more AT domains specific for ethylmalonyl CoA.
- 10 Illustrative PKS enzymes of this type include the FK-520 PKS and a recombinant PKS in which one or more AT domains is specific for ethylmalonyl CoA.

In a related embodiment, the present invention provides *Streptomyces* host cells in which one or more of the ethylmalonyl or 2-hydroxymalonyl biosynthetic genes have been deleted by homologous recombination or rendered inactive by mutation. For example, deletion or inactivation of the *fkbG* gene can prevent formation of the methoxyl groups at C-13 and C-15 of FK-520 (or, in the corresponding FK-506 producing cell, FK-506), leading to the production of 13,15-didesmethoxy-13,15-dihydroxy-FK-520 (or, in the corresponding FK-506 producing cell, 13,15-didesmethoxy-13,15-dihydroxy-FK-506). If the *fkbG* gene product acts on 2-hydroxymalonyl and the resulting 2-methoxymaionyl substrate is required for incorporation by the PKS, the AT domains of modules 7 and 8 may bind malonyl CoA and methylmalonyl CoA. Such incorporation results in the production of a mixture of polyketides in which the methoxy groups at C-13 and C-15 of FK-520 (or FK-506) are replaced by either hydrogen or methyl.

This possibility of non-specific binding results from the construction of a hybrid PKS of the invention in which the AT domain of module 8 of the FK-520 PKS replaced the AT domain of module 6 of DEBS. The resulting PKS produced, in *Streptomyces lividans*, 6-dEB and 2-desmethyl-6-dEB, indicating that the AT domain of module 8 of the FK-520 PKS could bind malonyl CoA and methylmalonyl CoA substrates. Thus, one could possibly also prepare the 13,15-didesmethoxy-FK-520 and corresponding FK-506 compounds of the invention by deleting or otherwise inactivating one or more or all of the genes required for 2-hydroxymalonyl CoA biosynthesis, i.e., the *fkbH*, *fkbI*, *fkbJ*, and *fkbK* genes. In any event, the deletion or inactivation of one or more biosynthetic genes required for ethylmalonyl and/or 2-hydroxymalonyl production prevents the formation of polyketides requiring ethylmalonyl and/or 2-hydroxymalonyl for biosynthesis, and the

resulting host cells are thus preferred for production of polyketides that do not require the same.

The host cells of the invention can be grown and fermented under conditions known in the art for other purposes to produce the compounds of the invention. See, e.g., 5 U.S. Patent Nos. 5,194,378; 5,116,756; and 5,494,820, incorporated herein by reference, for suitable fermentation processes. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. Preferred compounds of the invention include the following compounds: 13-desmethoxy-FK-506; 13-desmethoxy-FK-520; 13,15-didesmethoxy-FK-506; 13,15-didesmethoxy-FK-520; 13-desmethoxy-18-hydroxy-FK-506; 13-desmethoxy-18-hydroxy-FK-520; 13,15-didesmethoxy-18-hydroxy-FK-506; and 13,15-didesmethoxy-18-hydroxy-FK-520. These compounds can be further modified as described for tacrolimus and FK-520 in U.S. Patent Nos. 5,225,403; 5,189,042; 5,164,495; 5,068,323; 10 4,980,466; and 4,920,218, incorporated herein by reference.

15 Other compounds of the invention are shown in Figure 8, Parts A and B. In Figure 8, Part A, illustrative C-32-substituted compounds of the invention are shown in two columns under the heading R. The substituted compounds are preferred for topical administration and are applied to the dermis for treatment of conditions such as psoriasis. In Figure 8, Part B, illustrative reaction schemes for making the compounds shown in 20 Figure 8, Part A, are provided. In the upper scheme in Figure 8, Part B, the C-32 substitution is a tetrazole moiety, illustrative of the groups shown in the left column under R in Figure 8, Part A. In the lower scheme in Figure 8, Part B, the C-32 substitution is a disubstituted amino group, where R₃ and R₄ can be any group similar to the illustrative groups shown attached to the amine in the right column under R in Figure 25 8, Part A. While Figure 8 shows the C-32-substituted compounds in which the C-15-methoxy is present, the invention includes these C-32-substituted compounds in which C-15 is ethyl, methyl, or hydrogen. Also, while C-21 is shown as substituted with ethyl or allyl, the compounds of the invention includes the C-32-substituted compounds in which C-21 is substituted with hydrogen or methyl.

30 To make these C-32-substituted compounds, Figure 8, Part B, provides illustrative reaction schemes. Thus, a selective reaction of the starting compound (see Figure 8, Part B, for an illustrative starting compound) with trifluoromethanesulfonic anhydride in the presence of a base yields the C-32 O-triflate derivative, as shown in the upper scheme of Figure 8, Part B. Displacement of the triflate with 1H-tetrazole or 35 triazole derivatives provides the C-32 tetrazole or triazole derivative. As shown in the

lower scheme of Figure 8, Part B, reacting the starting compound with p-nitrophenylchloroformate yields the corresponding carbonate, which, upon displacement with an amino compound, provides the corresponding carbamate derivative.

The compounds can be readily formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation contains one or more of the compounds of the invention as an active ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use. Suitable formulation processes and compositions for the compounds of the present invention are described with respect to tacrolimus in U.S. Patent Nos. 5,939,427; 5,922,729; 5,385,907; 5,338,684; and 5,260,301, incorporated herein by reference. Many of the compounds of the invention contain one or more chiral centers, and all of the stereoisomers are included within the scope of the invention, as pure compounds as well as mixtures of stereoisomers. Thus the compounds of the invention may be supplied as a mixture of stereoisomers in any proportion.

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XIX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases relating to immunosuppression or neuronal damage, a compound of the invention may be administered orally, topically,

parenterally, by inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

5 Dosage levels of the compounds of the present invention are of the order from about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the
10 present invention may be administered on an intermittent basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral
15 administration to humans may contain from 0.5 mg to 5 g of active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention can be formulated within the range of,
20 for example, 0.00001% to 60% by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight. The compounds and compositions of the invention are useful in treating disease conditions using doses and administration schedules as described for tacrolimus in U.S. Patent Nos. 5,542,436; 5,365,948; 5,348,966; and 5,196,437, incorporated herein by reference. The compounds
25 of the invention can be used as single therapeutic agents or in combination with other therapeutic agents. Drugs that can be usefully combined with compounds of the invention include one or more immunosuppressant agents such as rapamycin, cyclosporin A, FK-506, or one or more neurotrophic agents.

It will be understood, however, that the specific dosage level for any particular
30 patient will depend on a variety of factors. These factors include the activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the present invention and shall not be construed as being a limitation on the scope of the invention or claims.

5

Example 1Replacement of Methoxyl with Hydrogen or Methyl at C-13 of FK-520

The C-13 methoxyl group is introduced into FK-520 via an AT domain in extender module 8 of the PKS that is specific for hydroxymalonyl and by methylation of the hydroxyl group by an S-adenosyl methionine (SAM) dependent methyltransferase.

10 Metabolism of FK-506 and FK-520 primarily involves oxidation at the C-13 position into an inactive derivative that is further degraded by host P450 and other enzymes. The present invention provides compounds related in structure to FK-506 and FK-520 that do not contain the C-13 methoxy group and exhibit greater stability and a longer half-life *in vivo*. These compounds are useful medicaments due to their immunosuppressive and

15 neurotrophic activities, and the invention provides the compounds in purified form and as pharmaceutical compositions.

20

The present invention also provides the novel PKS enzymes that produce these novel compounds as well as the expression vectors and host cells that produce the novel PKS enzymes. The novel PKS enzymes include, among others, those that contain an AT domain specific for either malonyl CoA or methylmalonyl CoA in module 8 of the FK-506 and FK-520 PKS. This example describes the construction of recombinant DNA compounds that encode the novel FK-520 PKS enzymes and the transformation of host cells with those recombinant DNA compounds to produce the novel PKS enzymes and the polyketides produced thereby.

25

To construct an expression cassette for performing module 8 AT domain replacements in the FK-520 PKS, a 4.6 kb *Sph*I fragment from the FK-520 gene cluster was cloned into plasmid pLitmus 38 (a cloning vector available from New England Biolabs). The 4.6 kb *Sph*I fragment, which encodes the ACP domain of module 7 followed by module 8 through the KR domain, was isolated from an agarose gel after 30 digesting the cosmid pKOS65-C31 with *Sph* I. The clone having the insert oriented so the single *Sac*I site was nearest to the *Spe*I end of the polylinker was identified and designated as plasmid pKOS60-21-67. To generate appropriate cloning sites, two linkers were ligated sequentially as follows. First, a linker was ligated between the *Spe*I and *Sac*I sites to introduce a *Bgl*II site at the 5' end of the cassette, to eliminate interfering polylinker sites, and to reduce the total insert size to 4.5 kb (the limit of the phage

35

KC515). The ligation reactions contained 5 picomolar unphosphorylated linker DNA and 0.1 picomolar vector DNA, i.e., a 50-fold molar excess of linker to vector. The linker had the following sequence:

5 5'-CTAGTGGGCAGATCTGGCAGCT-3'
3'-ACCCGTCTAGACCG-5'

The resulting plasmid was designated pKOS60-27-1.

Next, a linker of the following sequence was ligated between the unique *Sph*I and *Af*II sites of plasmid pKOS60-27-1 to introduce an *Nsi*I site at the 3' end of the module 8 cassette. The linker employed was:

10 5'-GGGATGCATGGC-3'
3'-GTACCCCTACGTACCGAATT-5'

The resulting plasmid was designated pKOS60-29-55.

To allow in-frame insertions of alternative AT domains, sites were engineered at the 5' end (*Avr* II or *Nhe* I) and 3' end (*Xho* I) of the AT domain using the polymerase chain reaction (PCR) as follows. Plasmid pKOS60-29-55 was used as a template for the PCR and sequence 5' to the AT domain was amplified with the primers SpeBgl-fwd and either *Avr*-rev or *Nhe*-rev:

SpeBgl-fwd 5'-CGACTCACTAGTGGGCAGATCTGG-3'
Avr-rev 5'-CACGCCTAGGCCGGTCGGTCTCGGGCCAC-3'
20 *Nhe*-rev 5'-GCGGCTAGCTGCTCGCCCATCGCGGGATGC-3'

The PCR included, in a 50 µl reaction, 5 µl of 10x *Pfu* polymerase buffer (Stratagene), 5 µl 10x z-dNTP mixture (2 mM dATP, 2 mM dCTP, 2 mM dTTP, 1 mM dGTP, 1 mM 7-deaza-GTP), 5 µl DMSO, 2 µl of each primer (10 µM), 1 µl of template DNA (0.1 µg/µl), and 1 µl of cloned *Pfu* polymerase (Stratagene). The PCR conditions were 95°C for 2 min., 25 cycles at 95°C for 30 sec., 60°C for 30 sec., and 72°C for 4 min., followed by 4 min. at 72°C and a hold at 0°C. The amplified DNA products and the Litmus vectors were cut with the appropriate restriction enzymes (*Bgl*II and *Avr*II or *Spe*I and *Nhe*I), and cloned into either pLitmus 28 or pLitmus38 (New England BioLabs), respectively, to generate the constructs designated pKOS60-37-4 and pKOS60-37-2, respectively.

30 Plasmid pKOS60-29-55 was again used as a template for PCR to amplify sequence 3' to the AT domain using the primers BsrXho-fwd and *Nsi*Afl-rev:

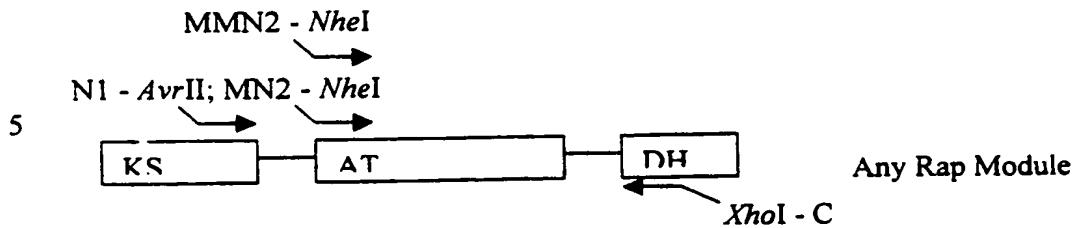
BsrXho-fwd 5'-GATGTACAGCTCGAGTCGGCACGCCGGCCGCATC-3'
35 *Nsi*Afl-rev 5'-CGACTCACTTAAGCCATGCATCC-3'

PCR conditions were as described above. The PCR fragment was cut with *Bsr*GI and *Af*II, gel isolated, and ligated into pKOS60-37-4 cut with *Asp*718 and *Af*II and

inserted into pKOS60-37-2 cut with *Bsr*GI and *Af*II, to give the plasmids pKOS60-39-1 and pKOS60-39-13, respectively. These two plasmids can be digested with *Avr*II and *Xba*I or *Nhe*I and *Xba*I, respectively, to insert heterologous AT domains specific for malonyl, methylmalonyl, ethylmalonyl, or other extender units.

5 Malonyl and methylmalonyl-specific AT domains were cloned from the rapamycin cluster using PCR amplification with a pair of primers that introduce an *Avr*II or *Nhe*I site at the 5' end and an *Xba*I site at the 3' end. The PCR conditions were as given above and the primer sequences were as follows:

- 10 RATN1 5'-ATCCTAGGCAGGCRGGYGTGTCGTCCCTCGG-3'
(3' end of Rap KS sequence and universal for malonyl and methylmalonyl CoA),
RATMN2 5'-ATGCTAGCCGCCGTTCCCGTCTCGCGCG-3'
(Rap AT shorter version 5'- sequence and specific for malonyl CoA),
RATMMN2 5'-ATGCTAGCGGATTCGTCGGTGGTGGTCCCGA-3'
(Rap AT shorter version 5'- sequence and specific for methylmalonyl CoA), and
15 RATC 5'-ATCTCGAGCCAGTASCGCTGGTGYTGGAAAGG-3'
(Rap DH 5'- sequence and universal for malonyl and methylmalonyl CoA).



10 Because of the high sequence similarity in each module of the rapamycin cluster, each primer was expected to prime any of the AT domains. PCR products representing ATs specific for malonyl or methylmalonyl extenders were identified by sequencing individual cloned PCR products. Sequencing also confirmed that the chosen clones contained no cloning artifacts. Examples of hybrid modules with the rapamycin AT12 and AT13 domains are shown in a separate figure.

15 The *AvrII-Xhol* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 12 of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below. The AT of rap module 12 is specific for incorporation of malonyl units.

20 AGATCTGGCAGCTGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 I W Q L A E A L L T L V R E S T
 GCCGCGGTGCTCGGCCACGTGGTGGCCAGGACATCCCCGCGACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTACCGCGGTCCAGCTGCGAACG 150
 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGGCGACCGGTGTGGCTGAACGCCACGGCGGTCTTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCACCCCGCACGTGCTCGCCGGGAAGCTCGGCGACGAAGTGGCGG 250
 F P T P H V L A G K L G D E L T G
 CACCCGCGCGCCCGTCGTGCCCGGACCGCGGCCACGGCGGTGCGCACG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGCTGCCGGGGGTGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCGCGGCTGGACGTCGACCGCATACGACCC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCCGACCGCATGGCAAGACCTTCGTCGCCACGGTGGCTTCCTC 500
 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGCTTCGACGCCGGCTTCTCGGACATCAGCCCGCGA 550
 T G A T G F D A A F F G I S P R E
 GGCCCTCGCGATGGACCCCGACAGCGGGTGTCTGGAGACGTGCGTGGG 600
 A L A M D P Q Q R V L L E T S W
 AGGGCGTTGAAAGCGCCGGCATACCCCGACTCGACCCGCGGAGCGAC 650
 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCCTTCTCTACGGTTACGGCACCGGGTGC 700
 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGACCGGGCTCGCAGACCAAGTGTGCTCTCCGGC 750
 T D G F G A T G S Q T S V L S G
 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCGACACG 800
 R L S Y F Y G L E G P A V T V D T
 GCGTGGTTCGTCGCTGGCGCTGCACCAAGGCCGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R

CTCCGGCGAATGCTCGCTGCCCTGGTCGGCGGTACCGGTATGGCGT 900
 S G E C S L A L V G G V T V M A
 CTCCCGCGGCTTCGTGGAGTCTCCCGGCAGCGCGGCCCTCGCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 5 GGCGGGCGAAGGCCTCGGCGCGGGTGCAGGACGGCACGAGCTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 GGGTCCGGTGTGCTGATCGTAGAGGGCTCTCGACGCCGAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTCACACCGTCCCTGGCGGTGCTGGTCCGGTCAACCAGGATGGT 1100
 10 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGCTGTCGGCGCCGAACGGGCCGTCAGGAGCAGGGTGAT 1150
 A S N G L S A P N G P S Q E R V I
 CCGGCAGGCCCTGGCCAACGCCGGCTCACCCCGCCGAGCTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 15 TCGAGGCCACGGCACCCGACAGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGGCCACCTACGGACAGGAGCGCCACCCCCCTGCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATGGCCACGCCAGGCCGCGTCCGGCGTCGCCG 1350
 20 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTCAGGCCCTCCGGCACGGGAGCTGCCGCCGACG 1400
 G I I K M V Q A L R H G E L P P T
 CTGCACGCCACGCCGTCGCCGACGTCGACTGGACGCCGGCGCGT 1450
 L H A D E P S P H V D W T A G A V
 25 CGAACTGCTGACGTGGCCCGCCGTGGCCCGAGACCGACGCCCTAGGC 1500
 E L L T S A R P W P E T D R P R
 GGGCAGGGCGTGTGTCCTTCGGGATCAGTGGCACCAACGCCACGTCACTC 1550
 R A G V S S F G I S G T N A H V I
 CTGGAAAGCGCACCCCCCACTCAGCCTGGGACAACGCCGCGTACGAGCG 1600
 30 L E S A P P T Q P A D N A V I E R
 GGCACCGGAGTGGGTGCCGTTGGTATCTGGCCAGGACCCAGTCGGCTT 1650
 A P E W V P L V I S A R T Q S A
 TGACTGAGCACGAGGGCCGGTGCCTGCGTATCTGGCGCGTCGCCGGG 1700
 L T E H E G R L R A Y L A A S P G
 35 GTGGATATGCCGGCTGTGGCATCGACGCTGGCGATGACACGGTCGGT 1750
 V D M R A V A S T L A M T R S V F
 CGAGCACCGTGCCTGCTGGGAGATGACACCGTCACCGGACCGCTG 1800
 E H R A V L L G D D T V T G T A
 TGTCTGACCCCTCGGGCGGTGTCGTCTTCCCGGGACAGGGGTCGCAGCGT 1850
 40 V S D P R A V F V F P G Q G S Q R
 GCTGGCATGGGTAGGAACTGGCCGCCGTTCCCGTCTCGCGCGGAT 1900
 A G M G E E L A A A F P V F A R I
 CCATCAGCAGGTGTGGACCTGCTCGATGTGCCGATCTGGAGGTGAACG 1950
 H Q Q V W D L L D V P D L E V N
 45 AGACCGGTTACGCCAGCCGGCCCTGTCGCAATGCAGGTGGCTCTGTC 2000
 E T G Y A Q P A L F A M Q V A L F
 GGGCTGCTGGAATCGTGGGGTGTACGACCGGACGCCGCGTACGCCATTG 2050
 G L L E S W G V R P D A V I G H S
 GGTGGGTGAGCTGGCTCGTATGTGTCGGGGTGTGGTCGGT 2100
 50 V G E L A A A Y V S G V W S L E
 ATGCCTGCACCTGGTGTGGCGCGGCTCGTCTGATGACGGCTCTGCC 2150
 D A C T L V S A R A R A L M Q A L P
 GCGGGTGGGTGATGGTCGCTGCTCCGGTCTGGAGGGATGAGGCCGGC 2200
 A G G V M V A V P V S E D E A R A
 55 CGTGCTGGGTGAGGGTGTGGAGATGCCGCCGTCAACGCCCGTCGG 2250
 V L G E G V E I A A V N G P S S
 TGGTTCTCCGGTGATGAGGCCGCCGTGCTGCAGGCCGCCGGCTG 2300
 V V L S G D E A A V L Q A A E G L
 GGGAGTGGACGCCGCTGGGACCCAGCCACGCCGTTCCATTCCGCCGTAT 2350
 60 G K W T R L A T S H A F H S A R M
 GGAACCCATGCTGGAGGAGTCCGGCGGTGCGCCGAAGGCCCTGACCTACC 2400
 E P M L E E F R A V A E G L T Y
 GGACGCCGCAGGTCTCCATGCCGTTGGTATCAGGTGACCACCGCTGAG 2450
 R T P Q V S M A V G D Q V T T A E

TACTGGGTGCCGGCAGGTCCGGGACACGGTCCGGTTCCGGGAGCAGGTGGC 2500
 Y W V R Q V R D T V R F G E Q V A
 CTCGTACGAGGACGCCGTGTTCGTCGAGCTGGTGGCAGCGGTCACTGG 2550
 S Y E D A V F V E L G A D R S L
 5 CCCGCCTGGTCGACGGTGTCCGGATGCTGCACGGCACCACGAAATCCAG 2600
 A R L V D G V A M L H G D H E I Q
 GCCCGCATCGGCGCCCTGGCCACCTGTATGTCAACGGCGTACCGTCGA 2650
 A A I G A L A H L Y V N G V T V D
 CTGGCCCGCGCTCTGGCGATGCTCCGGCAACACGGTGTGGACCTTC 2700
 10 W P A L L G D A P A T R V L D L
 CGACATACGCCCTCCAGCACCGCGTACTGGCTCGAGTCGGCACGCCCG 2750
 P T Y A F Q H Q R Y W L E S A R P
 GCCCGCATCCGACGGGGCCACCCCGTGTGGCTCCGGTATGCCCTCGC 2800
 A A S D A G H P V L G S G I A L A
 15 CGGGTCGCCGGGCGGGTGTTCACGGGTTCCGTGCCGACCGGTGCGGACC 2850
 G S P G R V F T G S V P T G A D
 GCGCGGTGTTCGTCGCCGAGCTGGCGCTGGCGCCGGACGGTGCAC 2900
 R A V F V A E L A L A A A D A V D
 TGCGCCACGGTCGAGCGGCTCGACATGCCCTCCGTGCCGGCGGG 2950
 20 C A T V E R L D I A S V P G R P G
 CCATGGCCGGACGACCGTACAGACCTGGTCGACGAGCCGGGACGACG 3000
 H G R T T V Q T W V D E P A D D
 GCCGGCGCCGGTTCACCGTCACACCCGACCGGCGACGCCCGTGGACG 3050
 G R R R F T V H T R T G D A P W T
 25 CTGCACGCCGAGGGGGTGTGCCGCCCCATGGCACGGCCCTGCCGATGC 3100
 L H A E G V L R P H G T A L P D A
 GGCGACGCCGAGTGGCCCCCACCGGGCGCGTGCACGGCGGACGGCTGC 3150
 A D A E W P P P G A V P A D G L
 CGGGTGTGGCGCCGGGGGACCAAGGTCTCGCCAGGCCGAGGTGGAC 3200
 30 P G V W R R G D Q V F A E A E V D
 GGACCGGACGGTTCTGGTCACCCGACCTGCTCGACGCCGTCTCTC 3250
 G P D G F V V H P D L L D A V F S
 CGCGGTCCGCGACGGAAGCCGCCAGCCGGGATGGCGCGACCTGACGG 3300
 A V G D G S R Q P A G W R D L T
 35 TGCACCGCTCGGACGCCACCGTACTGCCGCTGCCCTACCCGGCGACC 3350
 V H A S D A T V L R A C L T R R T
 GACGGAGCCATGGGATTGCCGCTTCGACGGCGCCGGCTGCCGGTACT 3400
 D G A M G F A A F D G A G L P V L
 CACCGCGAGGCGGTGACGCTGCCGGAGGTGGCGTCACCGTCGGCTCG 3450
 40 T A E A V T L R E V A S P S G S
 AGGAGTCGGACGCCCTGCAACCGGTTGGAGTGGCTGCCGGTGCACGG 3500
 E E S D G L H R L E W L A V A E A
 GTCTACCGACGGTGCACCTGCCGAGGGACATGTCTGATCACCGCCGCCA 3550
 V Y D G D L P E G H V L I T A A H
 45 CCCCCGACGACCCCGAGGACATACCCACCCGCCACACCCGCCACCC 3600
 P D D P E D I P T R A H T R A T
 GCGTCTGACCGCCCTGCAACACCCACCTCACCAACCGGACACACCC 3650
 R V L T A L Q H H L T T T D H T L
 ATCGTCCACACCAACCGACCCGCCGGCGCCACCGTCACCCGCTCAC 3700
 50 I V H T T D P A G A T V T G L T
 CCGCACCGCCAGAACGAAACACCCACCGCATCCGCTCATCGAAACCG 3750
 R T A Q N E H P H R I R L I E T
 ACCACCCCCACACCCCTCCCCCTGGCCCAACTGCCACCCCTGACCAC 3800
 D H P H T P L P L A Q L A T L D H
 55 CCCCCACCTCCGCTCACCCACCCACCCCTCACCAACCCACCC 3850
 P H L R L T H H T L H H P H L T P
 CCTCCACACCAACCCACCCACCCACCCACCCCTCAACCCGAACACG 3900
 L H T T T P P T T P L N P E H
 CCATCATCATCACCGCGGCTCCGGCACCCCTCGCCGGCATCTGCCCG 3950
 60 A I I I T G G S G T L A G I L A R
 CACCTGAACCACCCACACTACCTCTCTCCCGCACCCACCCCGA 4000
 H L N H P H T Y L L S R T P P P D
 CGCCACCCCCGGCACCCACCTCCCTGCGACGTGGCGACCCCCACCAAC 4050
 A T P G T H L P C D V G D P H Q

TCGCCACCACCCCTACCCACATCCCCAACCCCTCACCGCCATCTTCCAC 4100
 L A T T L T H I P Q P L T A I F H
 ACCGCCGCCACCCCTCGACGACGGCATCCTCCACGCCCTCACCCCGACCG 4150
 T A A T L D D G I L H A L T P D R
 5 CCTCACCAACCGTCTCCACCCAAAGCCAACGCCGCCCTGGCACCTGCACC 4200
 L T T V L H P K A N A A W H L H
 ACCTCACCCAAAACCAACCCCTACCCACTTCGTCTACTCCAGCGCC 4250
 H L T Q N Q P L T H F V L Y S S A
 GCCGCCGTCTCGGCCAGCCCCGACAAGGAAACTACGCCGCCAACGC 4300
 10 A A V L G S P G Q G N Y A A A N A
 CTTCCCTCGACGCCCTCGCCACCCACCGCACACCCCTCGCCAACCCGCCA 4350
 F L D A L A T H R H T L G Q P A
 CCTCCATCGCCTGGGCATGTGGCACACCACAGCACCCCTCACCGGACAA 4400
 15 T S I A W G M W H T T S T L T G Q
 CTCGACGACGCCGACCGGGACCGCATCCGCCGCCGGTTCCCTCCCGAT 4450
 L D D A D R D R I R R G G F L P I
 CACGGACGACGAGGGCATGGGATGCAT
 T D D E G

20 The *AvrII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS
 with the endogenous AT domain replaced by the AT domain of module 13 (specific for
 methylmalonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the
 amino acid sequence shown below.

25 AGATCTGGCAGCTGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 GCCGCCGTGCTGGCCACGTGGTGGCGAGGACATCCCCGCACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGTCAACCGCGGTCCAGCTGCACG 150
 F K D L G I D S L T A V Q L R N
 30 CCCTCACCGAGGCACCGGTGTGGCCTGAACGCCACGGCGGTCTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTGCCGGAAAGCTCGGCACGAACGTACCGG 250
 F P T P H V L A G K L G D E L T G
 CACCCGCGCGCCGTGTCGTCCCCGGACCGCGGCCACGGCGGTGCGCACG 300
 35 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGGCTGCCCGGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 A S P E E L W H L V A S G T D A I
 40 CACGGAGTTCCCGACGGACCGCGGCTGGGACGTGACGGCATACGACC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCUCGACCGCATCGGCAAGACCTTCGTCCGGCACGGTGGCTTC 500
 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGGCTTCGACCGCGGCTTCTCGGATCAGCCCGCGCA 550
 45 T G A T G F D A A F F G I S P R E
 GGCCCTCGCGATGGACCCCGCACGAGCGGGTGTCTGGAGACCTCGTGGG 600
 A L A M D P Q O R V L L E T S W
 AGGCCTTCGAAAGGCCGGCATACCCCGGACTCGACCCGCCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 50 ACCGGCGTGTCTCGTCGGCGCTTCTCCTACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGACCGGGCTCGCAGACCAAGTGTGCTCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCACAGC 800
 55 R L S Y F Y G L E G P A V T V D T
 GCGTGTTCGTCGTCGCTGGTGGCGCTGCACCAAGGCCGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTCGCCCTGGTCGGCGGTACGGTGTGGCGT 900
 S G E C S L A L V G G V T V M A
 60 CTCCCGCGGCTTCGTGGAGTTCTCCGGCAGCGCGGCCCTCGCGCCGGAC 950

S P G G F V E F S R Q R G L A P D
 GGCCGGGCGAAGCGGTTGGCGCGGGTGCGGACGGCACGAGCTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 GGGTGCCGGTGTGCTGATCGTCGAGAGGCTCTCCGACGCCAACGCAACG 1050
 5 G A G V L I V E R L S D A E R N
 GTCACACCGTCTGGCGTCGTCCTGGTGGCTCGCGGTCAACCAGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGGCTGCGGCCAACGGGCCGTCGAGGAGCGGGGTGAT 1150
 A S N G L S A P N G P S Q E R V I
 10 CCGGCAGGCCCTGGCCAACGCCGGCTCACCCCGGCCGGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 TCGAGGCCACGGCACCGGACCAGGCTGGCGACCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGCCACCTACGGACAGGAGCGCCACCCCCCTGCTGCTGGG 1300
 15 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAACGGCCGCGCCGTCGCCG 1350
 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTCAGGCCCTCCGGCACGGGGAGCTGCCGCCGACG 1400
 G I I K M V Q A L R H G E L P P T
 20 CTGCACGCCACGGCCGTCGCCGACGTCGACTGGACGCCGGCGCCGT 1450
 L H A D E P S P H V D W T A G A V
 CGAACTGCTGACGTGGCCGGCGTGGCCGAGACCGACCCGGCTAGGC 1500
 E L L T S A R P W P E T D R P R
 GGGCGGGCGTGTGTCCTCGGAGTCAGCGCACCAACGCCACGTCATC 1550
 25 R A G V S S F G V S G T N A H V I
 CTGGAGAGCGCACCCCCCGCTAGCCCGCGAGGGCAGCCCTGTTGA 1600
 L E S A P P A Q P A E E E A Q P V E
 GACGCCGGTGGTGGCTCGGATGTGCTGCCGCTGGTGAATATCGGCCAAGA 1650
 T P V V A S D V L P L V I S A K
 30 CCCAGCCGCCCTGACCGAACACGAAGACCGGCTGCCCTACTGGCG 1700
 T Q P A L T E H E D R L R A Y L A
 GCGTCGCCCGGGCGGATATACGGGCTGTGGCATCGACGCTGGCGGTGAC 1750
 A S P G A D I R A V A S T L A V T
 ACGGTCGGTGGTCGAGCACCGGCCGTACTCTTGGAGATGACACCGTCA 1800
 35 R S V F E H R A V L L G D D T V
 CCGGCACCGGGTGACCGACCCCAGGATCGTTGTCTTCCCGGGCAG 1850
 T G T A V T D P R I V F V F P G Q
 GGGTGGCAGTGGCTGGGATGGGAGTCGACTGCGCGATTGTCGGTGGT 1900
 G W Q W L G M G S A L R D S S V V
 40 GTTCGCCGAGCGGATGCCGAGTGTGCGGGCGTTGCGCGAGTCGTGG 1950
 F A E R M A E C A A A L R E F V
 ACTGGGATCTGTCACGGTCTGGATGATCCGGCGGTGGACCGGGTT 2000
 D W D L F T V L D D P A V V D R V
 GATGTGGTCAGCCCGCTCTGGCGATGATGGTTCCCTGGCCCGGT 2050
 45 D V V Q P A S W A M M V S L A A V
 GTGGCAGGGCGCCGGTGTGCGGGCGGATGCGGTGATCGGCCATTGCGAGG 2100
 W Q A A G V R P D A V I G H S Q
 GTGAGATGCCGCAGCTTGTGCGGGGTGCGGTGTCAGCGCATGCC 2150
 G E I A A A C V A G A V S L R D A
 50 GCCCGGATCGTGACCTGCGCAGCCAGGCATGCCGGGCTGGCGGG 2200
 A R I V T L R S Q A I A R G L A G
 CCGGGGCGCATGGCATCCGTCGCCCTGCCCGCAGGATGTCGAGCTGG 2250
 R G A M A S V A L P A Q D V E L
 TCGACGGGCGCTGGATGCCGCCAACACGGGCCGCTCCACCGTGATC 2300
 55 V D G A W I A A H N G P A S T V I
 GCGGGCACCCCGGAAGCGGTGACCATGTCCTCACCGCTCATGAGGCACA 2350
 A G T P E A V D H V L T A H E A Q
 AGGGGTGCGGGTGCAGGGATCACCGTCGACTATGCCCGCACACCCGC 2400
 G V R V R R I T V D Y A S H T P
 60 ACGTCGAGCTGATCCCGACGAACTACTCGACATCACTAGCGACAGCAGC 2450
 H V E L I R D E L L D I T S D S S
 TCGCAGACCCCGCTGCGCCGTGCGACCGTGACGGACCGCACCTGGGT 2500
 S Q T P L V P W L S T V D G T W V
 CGACAGCCCGCTGGACGGGAGTACTGGTACCGAACCTGCGTAACCGG 2550

D S P L D G E Y W Y R N L R E P
 TCGGTTTCCACCCCGCCGTCAAGCAGTTGCAGGCCAGGGCGACACCGTG 2600
 V G F H P A V S Q L Q A Q G D T V
 5 TTCGTCGAGGTCAAGCGCCAGCCCCGGTGTGTTGCAGGGCATGGACGACGA 2650
 F V E V S A S P V L L Q A M D D D
 TGTCGTACCGTTGCCACGCTGCGTCGTGACGACGGCGACGCCACCCGGA 2700
 V V T V A T L R R D D G D A T R
 TGCTCACCCCCCTGGCACAGGCCTATGTCCACGGCGTCACCGTGACTGG 2750
 M L T A L A Q A Y V H G V T V D W
 10 CCCGCCATCCTCGGCACCACCAACCCGGTACTGGACCTCCGACCTA 2800
 P A I L G T T T R V L D L P T Y
 CGCCCTCCAACACCAGCGGTACTGGCTCGAGTCGGCACGCCGGCGCAT 2850
 A F Q H Q R Y W L E S A R P A A
 CCGACGCCGGCCACCCCGTGTGGCTCCGGTATGCCCTCGCCGGTCG 2900
 15 S D A G H P V L G S G I A L A G S
 CCGGGCCGGGTGTTCACGGGTTCCGTGCCGACCGGTGCGGACCGCGCGT 2950
 P G R V F T G S V P T G A D R A V
 GTTCGTCGCCGAGCTGGCGTGGCGCCGGACGCGCTCGACTGCGCCA 3000
 F V A E L A L A A A D A V D C A
 20 CGGTGAGCGGGCTCGACATGCCCTCGTGCCTGGCCGGCCGGCATGGC 3050
 T V E R L D I A S V P G R P G H G
 CGGACGACCGTACAGACCTGGGTCGACGAGCCGGCGGACGACGGCG 3100
 R T T V Q T W V D E P A D D D G R R
 CCGGTTCACCGTGCACACCCGACCGGGCGACGCCCGTGGACGCTGCACG 3150
 25 R F T V H T R T G D A P W T L H
 CCGAUGGGGGTGTGCGCCCCCATGGCACGGCCCTGCCGATGCGGCCGAC 3200
 A E G V L R P H G T A L P D A A D
 GCCGAGTGGCCCCCACCGGGCGCGGTGCCCGCGACGGCTGCCGGTGT 3250
 A E W P P P G A V P A D G L P G V
 30 GTGGCGCCGGGGGACCAAGGTCTCGCCGAGGCCGAGGTGGACGGACCGG 3300
 W R R G D Q V F A E A E V D G P
 ACGGTTTGTGGTGACCCCCGACCTGCTGACCGGTCTCTCCGGTC 3350
 D G F V V H P D L L D A V F S A V
 GCGACGGAAGCCGCCAGCCGGCGGATGGCGGACCTGACGGTGCACGC 3400
 35 G D G S R Q P A G W R D L T V H A
 GTCGGACGCCACCGTACTGCGCGCTGCCCTACCCGGCGACCGACGGAG 3450
 S D A T V L R A C L T R R T D G
 CCATGGGATTGCGCCCTCGACGGCGCCGGCTGCCGTACTCACCGCG 3500
 A M G F A A F D G A G L P V L T A
 40 GAGGCGGTGACGCTGGGGAGGTGGCGTCACCGTCCGGCTCCGAGGAGTC 3550
 E A V T L R E V A S P S G S E E S
 GGACGGCCTGACCGGGTGGAGTGGCTCGCGGTGCCGAGGCGGTACG 3600
 D G L H R L E W L A V A E A V Y
 ACGGTGACCTGCCCAGGGACATGTCCTGATCACGCCGCCACCCGAC 3650
 45 D G D L P E G H V L I T A A H P D
 GACCCCGAGGACATACCCACCCGCCACACCCGCCACCCGCGTCC 3700
 D P E D I P T R A H T R A T R V L
 GACCGCCCTGCAACACCCACCTCACCAACCGACCCACCATCGTCC 3750
 T A L Q H H L T T D H T L I V
 50 ACACCAACCCGACCCCGCCGGCGCCACCGTCACCGGCTCACCCGACC 3800
 H T T D P A G A T V T G L T R T
 GCCCAGAACGAAACACCCCCACCGCATCCGCTCATCGAAACCGACCC 3850
 A Q N E H P H R I R L I E T D H P
 CCACACCCCCCTCCCCCTGGCCCAACTCGCCACCCCTGACCCACCC 3900
 55 H T P L P L A Q L A T L D H P H
 TCCGGCTCACCCACCAACCCCTCACCAACCCCCACCTCACCCCTCCAC 3950
 L R L T H H T L H H P H L T P L H
 ACCACCAACCCACCCACCAACCCCCCTCAACCCCGAACACGCCATCAT 4000
 T T T P P T T T P L N P E H A I I
 60 CATCACCGGGCTCCGGCACCCCTCGCCGGATCCTCGCCGCCACCTGA 4050
 I T G G S G T L A G I L A R H L
 ACCACCCCCACACCTACCTCCTCTCCCGCACCCCACCCCCCGACGCCACC 4100
 N H P H T Y L L S R T P P P D A T
 CCCGGCACCCACCTCCCTGCGACGTCGGCGACCCCCACCAACTCGCCAC 4150

P G T H L P C D V G D P H Q L A T
 CACCCCTCACCCACATCCCCAACCCCTACCGCCATCTCCACACCGCC 4200
 T L T H I P Q P L T A I F H T A
 CCACCCCTCGACGACGGCATCCTCCACGCCCTCACCCCGACCGCC CACC 4250
 5 A T L D D G I L H A L T P D R L T
 ACCGTCCTCCACCCAAAGCCAACGCCGCTGGCACCTGCACCACCTCAC 4300
 T V L H P K A N A A W H L H H L T
 CCAAAACCAACCCCTACCCACTTCGTCCTCTACTCCAGCGCCGCCG 4350
 Q N Q P L T H F V L Y S S A A A
 10 TCCCTGGCAGCCCCGACAAGGAAACTACGCCGCCAACGCCCTCCTC 4400
 V L G S P G Q G N Y A A A N A A F L
 GACGCCCTCGCCACCCACCGCCACCCCTCGGCCAACCCGCCACCTCCAT 4450
 D A L A T H R H T L G Q P A T S I
 15 CGCCTGGGGCATGTGGCACACCAACAGCACCCCTCACGGACAACTCGACG 4500
 A W G M W H T T S T L T G Q L D
 ACGCCGACCGGGACCGCATCCGGCGGGTTCCCTCCGATACGGAC 4550
 D A D R D R I R R G G F L P I T D
 GACGAGGGCATGGGGATGCAT
 20 D E G

The *NheII-XhoI* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 12 (specific for malonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

25 AGATCTGGCAGCTCGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 GCGCCGTGCTCGGCCACGTGGTGGCGAGGACATCCCCCGCAGGGCGGC 100
 A A V L G H V G G E D I P A T A A
 30 GTTCAAGGACCTGGCATCGACTCGCTCACCGCGGTCCAGCTCGCAACG 150
 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGCAGCCGGTGTGCGGCTGAACGCCACGGCGGTCTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTCGCCGGGAAGCTCGGCCACGAACCTGACCGG 250
 F P T P H V L A G K L G D E L T G
 35 CACCCGCGCCCGTCGTGCCCGGACCGCGGCCACGGCGGTGCGCACG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGCTGCCGGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 40 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCGCGCTGGACGTCGACCGATCTACGACC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCCGACCGCATCGCAAGACCTTCGTCGGCACGGTGGCTTCCTC 500
 45 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGCTTCGACCGCGCTTCTCGGCATCAGCCCGCGCA 550
 T G A T G F D A A F F G I S P R E
 GGGCCTCGCGATGGACCCCGCAGCAGCGGGTGTGCTCTGGAGACGTGCGTGG 600
 A L A M D P Q Q R V L L E T S W
 50 AGGCGTTCGAAAGCGCCGGCATACCCCGGACTCGACCCCGGGCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCTTCTCCACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGACCGGCTCGCAGACCGAGCTGTGCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 55 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCACACG 800
 R L S Y F Y G L E G P A V T V D T
 GCGTGTTCGTCGTCGGTGGCGCTGCACCAAGGCCGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTCGCCCTGGTCGGCGGTACGGTATGGCGT 900
 60 S G E C S L A L V G G V T V M A

CTCCCGGGCTTCGTGGAGTTCTCCGGCAGCGCAGGCCCTCGCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 GGCGGGGCAAGGCCTTCGGCGGGTGGGACGGCACGAGCTCGCCGA 1000
 5 G R A K A F G A G A D G T S F A E
 GGGTGCCTGGTGTGATCGTGGAGAGGGCTCTCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTCACACCGTCTGGCGGTGTCGGTGGTCTGGCGGTCAACCAGGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 10 GCCTCCAACGCCGTGTCGGCGCGAACGGGCCGTGCAAGGAGCGGGTGT 1150
 A S N G L S A P N G P S Q E R V I
 CGGGCAGGCCCTGGCCAACGCCGGCTACCCCGGCGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 TCGAGGCCACGGCACCGGACCCAGGCTGGCGACCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 15 GCGGTACTGGCACCTACGGACAGGAGCGCGCACCCCCCTGCTGCTGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAGGCCGTCCGGCGTCGCCG 1350
 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTCAGGCCCTCGGCCACGGGAGCTGCCGCCGACG 1400
 20 G I I K M V Q A L R H G E L P P T
 CTGCACGCCGACGCCGTGCGCACGTCGACTGGAGGCCGGCGCGT 1450
 L H A D E P S P H V D W T A G A V
 CGAACTGCTGACGTGGCCCGGGCTGGCCCAGACCGACCCGACGGC 1500
 E L L T S A R P W P E T D R P R
 25 GTGCCGCCGTCTCTCGTTGGGTGAGCGGACCAACGCCACGTCATC 1550
 R A A V S S F G V S G T N A H V I
 CTGGAGGCCGGACCGTAACGGAGACGCCCGGGCATGCCCTCCGGTGA 1600
 L E A C P V T E T P A A S P S G D
 CCTTCCCCCTGCTGGTGTGGCACGCTCACCGGAAGCGCTCGACGAGCAGA 1650
 30 L P L L V S A R S P E A L D E Q
 TCCGCCGACTGCGCGCCTACCTGGACACCACCCGGACGTCGACCGGGTG 1700
 I R R L R A Y L D T T P D V D R V
 GCCGTGGCACAGACGCTGGCCGGCGCACACACTTCGCCACCGCGCCGT 1750
 A V A Q T L A R R T H F A H R A V
 35 GCTGCTCGGTGACACCGTCATCACACACCCCCCGGGACCGGGCGACG 1800
 L L G D T V I T T P P A D R P D
 AACTCGTCTCGTCACTCCGCCAGGGCACCCAGCATCCCGATGGGC 1850
 E L V F V Y S G Q G T Q H P A M G
 GAGCAGCTAGCCGCCGTTCCCCGTCTCGCGCGATCCATCAGCAGGT 1900
 40 E Q L A A A F P V F A R I H Q Q V
 GTGGGACCTGCTCGATGTGCCGATCTGGAGGTGAACGAGACCGGTTACG 1950
 W D L L D V P D L E V N E T G Y
 CCCAGCCGCCCTGTCGAATGCAGGTGGCTCTGTTGGGCTGCTGGAA 2000
 45 A Q P A L F A M Q V A L F G L L E
 TCGTGGGGTGTACGCCGGACCGGGTGTGCGGCTATTGGTGGGTGAGCT 2050
 S W G V R P D A V I G H S V G E L
 TCGGGCTCGGTATGTGTCCGGGGTGTGGTCGTTGGAGGATGCCACTT 2100
 A A A Y V S G V W S L E D A C T
 TGGTGTGGCGCGGGCTCGTGTGAGCTGCGGCCATTGGTGGGTGGTG 2150
 50 L V S A R A R L M Q A L P A G G V
 ATGGTCGCTGTCCCCTCGGAGGATGAGGCCGGCGCTGCTGGGTGA 2200
 M V A V P V S E D E A R A V L G E
 GGGTGTGGAGATGCCCGGTCAACGGCCCGTCGTCGGTGGTTCTCTCG 2250
 G V E I A A V N G P S S V V L S
 55 GTGATGAGGCCGCCGTGCTGCAGGCCGCCAGGGCTGGGGAAAGTGGACG 2300
 G D E A A V L Q A A E G L G K W T
 CGGCTGGCGACCAGCCACGCCGTTCCATTCCGCCGTATGGAACCCATGCT 2350
 R L A T S H A F H S A R M E P M L
 GGAGGAGTCCGGCGGTGCCGAAGGCCGTGACCTACCGGACGCCGAGG 2400
 60 E E F R A V A E G L T Y R T P Q
 TCTCCATGCCGTTGGTGTGACCGTGACCACCGCTGAGTACTGGGTGCGG 2450
 V S M A V G D Q V T T A E Y W V R
 CAGGTCCGGGACACGGTCCGGTGGCGAGCAGGTGGCCTGACGAGGA 2500
 Q V R D T V R F G E Q V A S Y E D

CGCCGTTCGTCGAGCTGGGTGCCGACCGGTCACTGGCCCGCCTGGTCG 2550
 A V F V E L G A D R S L A R L V
 ACGGTGTCGCGATGTCGACGGCGACCACGAAATCCAGGCCGATCGGC 2600
 D G V A M L H G D H E I Q A A I G
 5 GCCCTGGCCCACCTGTATGTCAACGGCGTACGGTCGACTGGCCCGCCT 2650
 A L A H L Y V N G V T V D W P A L
 CCTGGCGATGCTCCGGAACACAGGGTCTGGACCTTCCGACATACGCCT 2700
 L G D A P A T R V L D L P T Y A
 TCCAGCACCGCGTACTGGCTCGAGTCGGCACGCCGCCGATCCGAC 2750
 10 F Q H Q R Y W L E S A R P A A S D
 GCAGGGCCACCCCGTGTGGCTCCGGTATCGCCCTGCCGGTCGCCGG 2800
 A G H P V L G S G I A L A G S P G
 CCGGGTGTTCACGGGTCCGTGCCGACCGGTGCGGACCGCGCGTGTTCG 2850
 R V F T G S V P T G A D R A V F
 15 TCGCGAGCTGGCGCTGGCCGCGCGACGGTCGACTGCGCACGGTC 2900
 V A E L A L A A A D A V D C A T V
 GAGCGGCTCGACATGCCCTCCGTGCCCGGCCGGCATGCCGGAC 2950
 E R L D I A S V P G R P G H G R T
 20 GACCGTACAGACCTGGGTGACGAGCCGGGACGACGGCCGGCGGT 3000
 T V Q T W V D E P A D D G R R R
 TCACCGTGACACCCCGACCGGGGACGCCCGTGGACGCTGACGCCGAG 3050
 F T V H T R T G D A P W T L H A E
 GGGGTGCTCGCCCCCATGGCACGGCCCTGCCGATGCCGGACGCCGA 3100
 G V L R P H G T A L P D A A D A E
 25 GTGGCCCCCACCGGGCGCGTGGCCCGGACGGCTGCCGGGTGTGGC 3150
 W P P P G A V P A D G L P G V W
 GCCGGGGGACCAAGGTCTTCGCCGAGGCCAGGTGGACGGACGGACGGT 3200
 R R G D Q V F A E A E V D G P D G
 TTCTGGTGCAACCCGACCTGCTCGACGCCGTCTCTCCCGGGTCGGCGA 3250
 30 F V V H P D L L D A V F S A V G D
 CGGAAGCCGCCAGCCGGCGGATGGCGCGACCTGACGGGTGACCGCTCGG 3300
 G S R Q P A G W R D L T V H A S
 ACGCCACCGTACTGCCGCCTGCCAACCCGGCGCACCGACGGAGCCATG 3350
 D A T V L R A C L T R R T D G A M
 35 GGATTGCCGCCTCGACGGCGCCGGCTGCCGGTACTCACCGCGGAGGC 3400
 G F A A F D G A G L P V L T A E A
 GGTGACGCTGCCGGAGGTGGCGTACCGTCCGGCTCCGAGGAGTCGGACG 3450
 V T L R E V A S P S G S E E S D
 GCCTGCACCGGTTGGAGTGGCTCGCGGTGCCGAGGCCGGTACCGACGGT 3500
 40 G L H R L E W L A V A E A V Y D G
 GACCTGCCGAGGGACATGTCCCTGATCACCGCCGCCACCCGACGACCC 3550
 D L P E G H V L I T A A H P D D P
 CGAGGACATACCCACCCCGCGCCACACCCCGCGCACCCGCGTCTGACCG 3600
 E D I P T R A H T R A T R V L T
 45 CCCTGCAACACCAACCTCACCAACCGACCCACACCCCTCATGTCCACACC 3650
 A L Q H H L T T D H T L I V H T
 ACCACCGACCCCGCCGGCGCACCGTCACCGGCCCTCACCGCACCGCCCA 3700
 T T D P A G A T V T G L T R T A Q
 50 GAACGAACACCCCCACCGCATCCGCCTCATCGAAACCGACCCACCA 3750
 N E H P H R I R L I E T D H P H
 CCCCCCTCCCCCTGGCCCAACTGCCACCCCTGACCAACCCACCTCCGC 3800
 T P L P L A Q L A T L D H P H L R
 CTCACCCACCAACACCCCTCACCCACCCACCTCACCCCCCTCCACACCAC 3850
 L T H H T L H H P H L T P L H T T
 55 CACCCACCCACCAACCCCCCTCAACCCGAACACGCCATCATCA 3900
 T P P T T T P L N P E H A I I I
 CCGGCGGCTCCGGCACCCCTGCCGGCATCCCTGCCGCCACCTGAACAC 3950
 T G G S G T L A G I L A R H L N H
 CCCCCACACCTACCTCCCTCCCGCACCCACCCCCCGACGCCACCCCGG 4000
 60 P H T Y L L S R T P P P D A T P G
 CACCCACCTCCCCCTGCCGACGTGGCGACCCCCCACCAACTGCCACCCAC 4050
 T H L P C D V G D P H Q L A T T
 TCACCCACATCCCCAACCCCTCACCGCCATCTCCACACCACGCCACC 4100
 L T H I P Q P L T A I F H T A A T

CTCGACGACGGCATCCTCCACGCCCTCACCCCCGACCGCCTCACCCACCGT 4150
 L D D G I L H A L T P D R L T T V
 CCTCCACCCCAAAGCCAACGCCGCTGGCACCTGCACCCACCTCACCCAAA 4200
 L H P K A N A A W H L H L T Q
 5 ACCAACCCCTCACCCACTTCGTCTCTACTCCAGCGCCGCCGCGTCCTC 4250
 N Q P L T H F V L Y S S A A A V L
 GGCAGCCCCGGACAAGGAAACTACGCCGCCAACGCCCTCGACGC 4300
 G S P G Q G N Y A A A N A F L D A
 CCTCGCCACCCACCAGCCACACCTCGGCCAACCCGCCACCTCCATCGCCT 4350
 10 L A T H R H T L G Q P A T S I A
 GGGGCATGTGGCACACCACCGACCCCTCACCGGACAACTCGACGACGCC 4400
 W G N W H T T S T L T G Q L D D A
 GACCGGGACCGCATCGCCGCCGGTTCCCTCCCGATCACGGACGACGA 4450
 D R D R I R R G G F L P I T D D E
 15 GGGCATGGGGATGCAT
 G

The *NheII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS
 with the endogenous AT domain replaced by the AT domain of module 13 (specific for
 20 methylmalonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the
 amino acid sequence shown below.

AGATCTGGCAGCTGCCGAAGCGCTGCTGACGCTCGTCCGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 25 GCGCCCGTGTCTGGCACGTGGTGGCGAGGACATCCCCGCACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTCACCGCGGTCCAGCTGCGCAACG 150
 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGCAGCCGGTGTGGCTGAACGCCACGGCGGTCTCGAC 200
 A L T E A T G V R L N A T A V F D
 30 TTCCCGACCCGCACGTGCTGCCGGAAAGCTCGGCACGAACGTACCGG 250
 F P T P H V L A G K L G D E L T G
 CACCCGCGCGCCCGTGTGGCCCGGACCGCGGCCACGGCGGTGCGCACG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGGCTGCCCGGGGTC 350
 35 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 A S P E E L W H L V A S G T D A I
 CACGGAGTCCCGACGGACCGCGGCTGGACGTCGACCGATCTACGACC 450
 T E F P T D R G W D V D A I Y D
 40 CGGACCCGACCGATCGGCAAGACCTTCGTCCGGCACGGTGGCTTCCTC 500
 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGCTTCGACGCCGTTCTCGGCATCAGCCCGCGCA 550
 T G A T G F D A A F F G I S P R E
 GGCCCTCGCGATGGACCCGCAGCAGCGGGTGCTCTGGAGACGTCGTGG 600
 45 A L A M D P Q Q R V L L E T S W
 AGGCCTTCGAAAGCGCCGGCATACCCCGGACTCGACCCGCCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 ACCGGCGTTCGTGGCGCCCTCTCTACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 50 CACCGACGGCTTCGGCGACCGGCTCGCAGACCGAGCTGTGCTCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCACACG 800
 R L S Y F Y G L E G P A V T V D T
 GCGTGTTCGTGCGCTGGTGGCGCTGCACCCAGGCGGGCAGTCGCTGCG 850
 55 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTCGCCCTGGTCGGCGCGTACGGTGATGGCGT 900
 S G E C S L A L V G G V T V M A
 CTCCCAGGCGGCTTCGTGGAGTTCTCCCGGAGCGCGGCCCTCGCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 60 GGCGGCTCGAAGGCCTTCGGCGGGTGCGGACGGCACGGAGCTCGCCGA 1000

G R A K A F G A G A D G T S F A E
 GGGTCCGGTGTGCTGATCGTCGAGAGGCCTCCGACGCCGAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 5 GTCACACCGTCCTGGCGGTGTCGGTGGTCGGCGGTCAACCAGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 GCCTCCAACGGGCTGTCGGCGCCGAACGGGCCGTCGAGGAGCGGGTGAT 1150
 A S N G L S A P N G P S Q E R V I
 CCGGCAGGCCCTGGCCAACGCCGGCTCACCCGGCGGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 10 TCGAGGCCACGGCACCCAGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGGCCACCTACGGACAGGAGCGCCACCCCCCTGCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCCAACATCGGCCACGCCAGGCCGCGTCCGGCGTCGCG 1350
 15 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGAGCTGCCGCCGACG 1400
 G I I K M V Q A L R H G E L P P T
 CTGCACGCCGACGCCGGTCCGGCACGTCGACTGGACGCCGCCGT 1450
 L H A D E P S P H V D W T A G A V
 20 CGAACTGCTGACGTCCGGCCGGCGTGGCCGAGACCGACGCCACGGC 1500
 E L L T S A R P W P E T D R P R
 GTGCCGCCGTCTCCTCGTTGGGGTGGCCGACCAACGCCACGTCACTC 1550
 R A A V S S F G V S G T N A H V I
 CTGGAGGCCGGACCGTAACGGAGACGCCGCCGACGCCCTCCGGTGA 1600
 25 L E A G P V T E T P A A S P S G D
 CCTTCCCCCTGCTGGTGTGGCACGCTCACCGGAAGCGCTGACGAGCAGA 1650
 L P L L V S A R S P E A L D E Q
 TCCGCCGACTGCGCGCTACCTGGACACCACCCGGACGTCGACCGGGTG 1700
 I R R L R A Y L D T T P D V D R V
 30 GCCGTGGCACAGACGCTGGCCGGCGCACACACTTCGCCAACCGCGCGT 1750
 A V A Q T L A R R T H F A H R A V
 GCTGCTGGTGCACCCGTACCAACACCCCCCGCCGACCGGGCCGACG 1800
 L L G D T V I T T P P A D R P D
 AACTCGTCTCGTCACTCCGGCAGGGCACCCAGCATCCCGCATGGGC 1850
 35 E L V F V Y S G Q G T Q H P A M G
 GAGCAGCTAGCCGATTGTCGGTGGTTCGCCGAGCGGATGGCGAGTG 1900
 E Q L A D S S V V F A E R M A E C
 TGCGGCCGGCGTGCACGGTGGACTGGGATCTGTTACGGTCTGG 1950
 A A A L R E F V D W D L F T V L
 40 ATGATCCGGCGGTGGTGGACCGGGTTGATGTTGTCAGCCGCTTCTGG 2000
 D D P A V V D R V D V V Q P A S W
 GCGATGATGGTTCCCTGGCCGGTGTGGCAGGCCGGGTGTGGCC 2050
 A M M V S L A A V W Q A A G V R P
 GGATGCGGTGATCGGCCATTGCAAGGGTGAGATGCCGCAGCTTGTGG 2100
 45 D A V I G H S Q G E I A A A C V
 CGGGTGCGGTGTCACTACCGCATGCCGCCGGATCGTGCACCTGGCGCAGC 2150
 A G A V S L R D A A R I V T L R S
 CAGGCATGCCGCCGGGCTGGCGGGCGCGATGGCATCCGTCGC 2200
 50 Q A I A R G L A G R G A M A S V A
 CCTGCCGCCGAGGATGTCGAGCTGGTCGACGGGCTGGATGCCGCC 2250
 L P A Q D V E L V D G A W I A A
 ACAACGGGCCGCCCTCCACCGTATGCCGCCGGATCGTGCACCTGGCGCAGC 2300
 H N G P A S T V I A G T P E A V D
 CATGTCCTCACCGCTATGAGGCACAAGGGTGCGGGTGCAGGCCGATCAC 2350
 55 H V L T A H E A Q G V R V R R I T
 CGTCGACTATGCCCTCGCACACCCGCACGTCGAGCTGATCCCGCACGAAC 2400
 V D Y A S H T P H V E L I R D E
 TACTCGACATCACTAGCGACAGCAGCTCGAGACCCCCGCTGTGGCGTGG 2450
 60 L L D I T S D S S S Q T P L V P W
 CTGTCGACCGTGGACGGCACCTGGTCGACAGCCCGCTGGACGGGGAGTA 2500
 L S T V D G T W V D S P L D G E Y
 CTGGTACCGGAACCTGCGTGAACCGGTGCGTTCCACCCGCCGTCAGCC 2550
 W Y R N L R E P V G F H P A V S
 AGTTGCAGGCCAGGGCAGACCCGTGTTCGTCAGGTCAGGCCAGCCCG 2600

Q L Q A Q G D T V F V E V S A S P
 GTGTTGTTGCAGGCAGGATGGACGACGATGTCGTCACGGTTGCCACGCTGCG 2650
 V L L Q A M D D D V V T V A T L R
 TCGTGACGACGGCAGGCCACCCGGATGCTCACCGCCCTGGCACAGGCCT 2700
 5 R D D G D A T R M L T A L A Q A
 ATGTCCACGGCGTCACCGTCGACTGGCCGCCATCTCGGCACCCACCA 2750
 Y V H G V T V D W P A I L G T T T
 ACCCGGGTACTGGACCTTCCGACCTACGCCTTCCAACACCAGGGTACTG 2800
 T R V L D L P T Y A F Q H Q R Y W
 10 GCTCGAGTCGGCACGCCCGGCCATCCGACGCCGGCACCCGGTCTGG 2850
 L E S A R P A A S D A G H P V L
 GCTCCGGTATGCCCTCGCCGGTCGCCGGGGTGTTCACGGGTTCC 2900
 G S G I A L A G S P G R V F T G S
 GTGCCGACCGGTGCCGACCGCGCGGTGTTCTCGTCGCCAGCTGGCCTGGC 2950
 15 V P T G A D R A V F V A E L A L A
 CGCCGCGGACGCCGGTCACTGCCACGGTCGAGCGGCTCGACATGCC 3000
 A A D A V D C A T V E R L D I A
 CCGTGCCCCGGCCGGCCGGCATGGCCGACCGTACAGACCTGGTC 3050
 20 S V P G R P G H G R T T V Q T W V
 GACGAGCCGGCGGACGACGGCCGGCGCCGGTTCACCGTGCACACCCGCAC 3100
 D E P A D D G R R R F T V H T R T
 CGGCGACGCCCGTGGACGCTGCACGCCGAGGGGGTGTGCGCCCCATG 3150
 G D A P W T L H A E G V L R P H
 GCACGGCCCTGCCUGATGCCGACGCCGAGTGGCCCCCACGGCGCG 3200
 25 G T A L P D A A D A E W P P P G A
 GTGCCCGGGACGGCTGCCGGGTGTGTGGCGCCGGGGGACCAAGGTCTT 3250
 V P A D G L P G V W R R G D Q V F
 CGCCGAGGGCCGAGGTGGACGGACCGGACGGTTCTGTGGTGCACCCGACC 3300
 A E A E V D G P D G F V V H P D
 30 TGCTGACCGCGGTCTCTCCGGTCCGGACGGAACGCCAGCCGGCC 3350
 L L D A V F S A V G D G S R Q P A
 GGATGGCGCGACCTGACGGTCAACCGCTGGACGCCACCGTACTGCCGC 3400
 G W R D L T V H A S D A T V L R A
 CTGCCTCACCCGGCGACCGACGGAGCCATGGGATTGCCGCCTCGACG 3450
 35 C L T R R T D G A M G F A A F D
 GCGCCGGCCCTGCCGGTACTCACCGCGAGGGCTGACGCTGCCGGAGGTG 3500
 G A G L P V L T A E A V T L R E V
 GCGTCACCGTCCGGCTCCGAGGAGTCGGACGCCCTGCACCGGTTGGAGTG 3550
 A S P S G S E E S D G L H R L E W
 40 GCTCGCGGTGCCGAGGCAGGTCTACGACGGTACCTGCCGAGGGACATG 3600
 L A V A E A V Y D G D L P E G H
 TCCCTGATCACCGCCGCCACCCCGACGACCCCGAGGACATACCCACCCGC 3650
 V L I T A A H P D D P E D I P T R
 GCCCACACCCCGGCCACCGCGCTCTGACCCCTGCAACACCACTCAC 3700
 45 A H T R A T R V L T A L Q H H L T
 CACCAACCGACCAACCCCTCATCGTCCACACCACCGACCCCGCCGGCG 3750
 T T D H T L I V H T T T D P A G
 CCACCGTCACCGGCTCACCGCACCGCCAGAACGAAACACCCACCGC 3800
 50 A T V T G L T R T A Q N E H P H R
 ATCCGCCTATCGAAACCGACCAACCCACACCCCTCCGGCCA 3850
 I R L I E T D H P H T P L P L A Q
 ACTCGCCACCCCTGACCAACCCACCTCCGGCTCACCCACCACTCC 3900
 L A T L D H P H L R L T H H T L
 ACCACCCACCTCACCCCTCCACACCACCCACCCACCCACC 3950
 55 H H P H L T P L H T T T P P T T T
 CCCCTCAACCCCGAACACGCCATCATCATCACCGGGCTCCGGCACCC 4000
 P L N P E H A I I I T G G S G T L
 CGCCGGCATCCCTGCCGCCACCTGAACCAACCCACACCTACCTCC 4050
 A G I L A R H L N H P H T Y L L
 60 CCGCACCCACCCCGACGCCACCCCGACCCACCTCCCGAC 4100
 S R T P P P D A T P G T H L P C D
 GTCGGCGACCCCCACCAACTGCCACCCACCTCACCCACATCCCCAAC 4150
 V G D P H Q L A T T L T H I P Q P
 CCTCACCGCCATCTCCACACCGCCGCCACCTCGACGACGGCATCCTCC 4200

```

L T A I F H T A A T L D D G I L
ACGCCCTCACCCCGACCGCCTCACCAACCGCTCCACCCAAAGCCAAC 4250
H A L T P D R L T T V L H P K A N
GCCGCCTGGCACCTGCACCACCTCACCCAAACCAACCCCTCACCCACTT 4300
5 A A W H L H H L T Q N Q P L T H F
CGTCCTCTACTCCAGCGCCGCCGTCCCTGGCAGCCCCGGACAAGGAA 4350
V L Y S S A A A V L G S P G Q G
ACTACGCCGCCAACGCCCTCGACGCCCTGCCACCCACCGCCAC 4400
N Y A A A A N A F L D A L A T H R H
10 ACCCTCGGCCAACCCGCCACCTCATCGCCTGGGCATGTGGCACACCAC 4450
T G Q P A T S I A W G M W H T T
CAGCACCCCTCACCGGACAACCTCGACGACGCCGACCGGGACCGCATCCGCC 4500
S T L T G Q L D D A D R D R I R
GCGGCGGTTTCCTCCCGATCACGGACGACGAGGGCATGGGATGCAT
15 R G G F L P I T D D E G

```

Phage KC515 DNA was prepared using the procedure described in *Genetic Manipulation of Streptomyces, A Laboratory Manual*, edited by D. Hopwood *et al.* A phage suspension prepared from 10 plates (100 mm) of confluent plaques of KC515 on *S. lividans* TK24 generally gave about 3 µg of phage DNA. The DNA was ligated to circularize at the cos site, subsequently digested with restriction enzymes *Bam*HI and *Pst*I, and dephosphorylated with SAP.

Each module 8 cassette described above was excised with restriction enzymes *Bgl*II and *Nsi*I and ligated into the compatible *Bam*HI and *Pst*I sites of KC515 phage DNA prepared as described above. The ligation mixture containing KC515 and various cassettes was transfected into protoplasts of *Streptomyces lividans* TK24 using the procedure described in *Genetic Manipulation of Streptomyces, A Laboratory Manual* edited by D. Hopwood *et al.* and overlaid with TK24 spores. After 16-24 hr, the plaques were restreaked on plates overlaid with TK24 spores. Single plaques were picked and resuspended in 200 µL of nutrient broth. Phage DNA was prepared by the boiling method (Hopwood *et al.*, *supra*). The PCR with primers spanning the left and right boundaries of the recombinant phage was used to verify the correct phage had been isolated. In most cases, at least 80% of the plaques contained the expected insert. To confirm the presence of the resistance marker (thiostrepton), a spot test is used, as described in Lomovskaya *et al.* (1997), in which a plate with spots of phage is overlaid with mixture of spores of TK24 and phiC31 TK24 lysogen. After overnight incubation, the plate is overlaid with antibiotic in soft agar. A working stock is made of all phage containing desired constructs.

Streptomyces hygroscopicus ATCC 14891 (see US Patent No. 3,244,592, issued 40 5 Apr 1966, incorporated herein by reference) mycelia were infected with the recombinant phage by mixing the spores and phage (1×10^8 of each), and incubating on R2YE agar (*Genetic Manipulation of Streptomyces, A Laboratory Manual*, edited by D.

Hopwood *et al.*) at 30°C for 10 days. Recombinant clones were selected and plated on minimal medium containing thiostrepton (50 µg/ml) to select for the thiostrepton resistance-conferring gene. Primary thiostrepton resistant clones were isolated and purified through a second round of single colony isolation, as necessary. To obtain 5 thiostrepton-sensitive revertants that underwent a second recombination event to evict the phage genome, primary recombinants were propagated in liquid media for two to three days in the absence of thiostrepton and then spread on agar medium without thiostrepton to obtain spores. Spores were plated to obtain about 50 colonies per plate, and thiostrepton sensitive colonies were identified by replica plating onto thiostrepton 10 containing agar medium. The PCR was used to determine which of the thiostrepton sensitive colonies reverted to the wild type (reversal of the initial integration event), and which contain the desired AT swap at module 8 in the ATCC 14891-derived cells. The PCR primers used amplified either the KS/AT junction or the AT/DH junction of the wild-type and the desired recombinant strains. Fermentation of the recombinant strains, 15 followed by isolation of the metabolites and analysis by LCMS, and NMR is used to characterize the novel polyketide compounds.

Example 2

Replacement of Methoxyl with Hydrogen or Methyl at C-13 of FK-506

20 The present invention also provides the 13-desmethoxy derivatives of FK-506 and the novel PKS enzymes that produce them. A variety of *Streptomyces* strains that produce FK-506 are known in the art, including *S. tsukubaensis* No. 9993 (FERM BP-927), described in U.S. Patent No. 5,624,852, incorporated herein by reference; *S. hygroscopicus* subsp. *yakushimaensis* No. 7238, described in U.S. patent No. 4,894,366, 25 incorporated herein by reference; *S. sp.* MA6858 (ATCC 55098), described in U.S. Patent Nos. 5,116,756, incorporated herein by reference; and *S. sp.* MA 6548, described in Motamedi *et al.*, 1998, "The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK-506," *Eur. J. Biochem.* 256: 528-534, and Motamedi *et al.*, 1997, "Structural organization of a multifunctional polyketide synthase involved in the 30 biosynthesis of the macrolide immunosuppressant FK-506," *Eur. J. Biochem.* 244: 74-80, each of which is incorporated herein by reference.

The complete sequence of the FK-506 gene cluster from *Streptomyces* sp. MA6548 is known, and the sequences of the corresponding gene clusters from other FK-506-producing organisms is highly homologous thereto. The novel FK-506 recombinant 35 gene clusters of the present invention differ from the naturally occurring gene clusters in

that the AT domain of module 8 of the naturally occurring PKSs is replaced by an AT domain specific for malonyl CoA or methylmalonyl CoA. These AT domain replacements are made at the DNA level, following the methodology described in Example 1.

5 The naturally occurring module 8 sequence for the MA6548 strain is shown below, followed by the illustrative hybrid module 8 sequences for the MA6548 strains.

10 GCATGCGGCTGTACCGAGGCGGCACGGCGCACCGGAAGTCCCCTGGTGGTG 50
 M R L Y E A A R R T G S P V V V
 GCGGCCGCGCTCGACGACGCCGGACGTGCCGCTGCTGCGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCGTCCGCCGTCGGGAACGCTCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCGCCGACGCCCTCCCTCGCGTTCG 200
 R S P C C P T T S A P T P P S R S
 TCCCTGGAACAGCACCGCCACCGTGTCCGGCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGAGCTCAAGGAACTCGGCATCGACTCGCTACCGCGG 300
 P A T T T F K E L G I D S L T A
 TCCAGCTGCGCAACCGCCTGACCAACGGGACCGGGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCGCCGCGCTCGCCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCGGCGCCCGTCGCGGCCCGGACCGCGGCA 450
 D E L A G T R A P V A A R T A A
 25 CCGCGGCCGCGCACGACGAACCGCTGGCGATCGTGGGCATGGCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGCGGGGTGCGTCGCCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATCACGGAGTTCCCCGGACCGCGGCTGGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACAGCGCTTACGACCCGGACCCCGACGCCGATCGGCAAGACCTTCGTCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCCTCGACGGTGCACGCCGCTCGACGCCGCGCTTCGG 700
 H G G F L D G A T G F D A A A F F G
 35 GATCAGCCCGCGAGGCCATGGACCCCGCAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCCTGGAGGCCTGAAAGCGCGGGCATACCCCGGACGCG 800
 L E T S W E A F E S A G I T P D A
 GCGCGGGCGAGCGACACCGCGTGTTCATCGGCCGTTCTCCACGGTA 850
 A R G S D T G V F I G A F S Y G Y
 40 CGGCACCGCGTGCCTGAAACACGGCTTCGGCGCGACAGGGTGCAGACCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGCCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 45 GTCACGGTCGACACCCGCTGCTCGTCGACTGGTCGCCCTGCACCGAGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTCGCCTCGGGCAATGCTCGCTGCCCTGGTCGGCGGTG 1050
 G Q S L R S G E C S L A L V G G
 TCACGGTGATGGCGTCGCCCGGGATTCTCGAGTTCTCCGGCAGCGC 1100
 50 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGCGGGCGAAGGGCTTCGGCGCGGGCGGACGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTGGTCGAGCGGCTCTCG 1200
 T S F A E G A G A L V V E R L S
 55 ACGCGGAGCGCCACGGCCACACCGTCTCGCCCTCGTACCGCGCTCCGCG 1250
 D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCCGAACGGCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCGTCATCCACCAAGGCCCTCGCGAACCGCGAAACTCACCCCG 1350

Q E R V I H Q A L A N A K L T P
 CCGATGTCGACGCCGTCGAGGCGCACGGCACCGGACCGCCTCGGCAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCGCAGGCGCTCGCGACGTACGGACAGGACCGGGCGAC 1450
 5 P I E A Q A L L A T Y G Q D R A T
 GCCCCCTGCTCGCTGGCTCGCTGAAGTCAACATCGGGCACGCCAGGCC 1500
 P L L L G S L K S N I G H A Q A
 CGTCAGGGCGCCGGATCATCAAGATGGTCAGGCCATCCGGCACGGG 1550
 10 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCACACTGCACGCCAGGAGCCGTCGCCACGTCACTG 1600
 E L P P T L H A D E P S P H V D W
 GACGGCCGGTGCCTCGAGCTCCTGACGTGGCCCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTGCCCGCGCCGCGCTCGCTCGTCTCGTGGCGTGGCGCACG 1700
 15 T G R P R R A A V S S F G V S G T
 AACGCCACATCATCCTGAGGCAGGACCGGTCAAAACGGGACCGTCGA 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCGATCGAGGCAGGACCGGTCAAGTAGGACCGGTCAAGGCTG 1800
 20 A G A I E A G P V E V G P V E A
 GACCGCTCCCCGCCGCCGCGCTCAGCACCGGGCGAAGACCTTCCGCTG 1850
 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCGTCCCCGGAGGCACTCGACGAGCAGATCGGGCGCCT 1900
 L V S A R S P E A L D E Q I G R L
 GCGGCCCTATCTGACACCGGCCGGCGTCGACCGGGCGGGCGTGGCGC 1950
 25 R A Y L D T G P G V D R A A V A
 AGACACTGGCCCGGTACGCACTTCACCCACCGGGCGTACTGCTCGGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTCATCGGCCTCCCCCGCGGACCAGGCCACGAACCTCGTCTT 2050
 D T V I G A P P A D Q A D E L V F
 30 CGTCTACTCCGGTCAGGGCACCCAGCATCCCGCATGGCGAGCAACTCG 2100
 V Y S G Q G T Q H P A M G E Q L
 CGGCCGCGTCCCCGTGTTGCCGATGCCACGCCGCGCTCCGACGG 2150
 A A A F P V F A D A W H D A L R R
 CTCGACGACCCCGACCCGACCGACCCACACGGAGGCCAGCACCGCTCTT 2200
 35 L D D P D P H D P T R S Q H T L F
 CGGCCACCCAGGGCGTTCACCGCCCTCTGAGGTCTGGACATCACGC 2250
 A H Q A A F T A L L R S W D I T
 CGCACGCCGTATCGGCCACTCGCTGGCGAGATACCGCCCGTACGCC 2300
 40 P H A V I G H S L G E I T A A Y A
 GCCGGGATCCTGTCGCTGACGACGCCGACCCCTGATCACCACCGCTGC 2350
 A G I L S L D D A C T L I T T R A
 CCGCCTCATGCACACGCTTCCGCCGCCGCGCCATGGTCACCGTCTGA 2400
 R L M H T L P P P G A M V T V L
 CCAGCGAGGAGGAGGAGGCCGTCAGGCCTGCGCTGCCGGCGTGGAGATGCC 2450
 45 T S E E E A R Q A L R P G V E I A
 GCGGTCTCGGCCGCACTCCGTCGTGCTCTGGCGACGAGGACGCCGT 2500
 A V F G P H S V V L S G D E D A V
 GCTCGACGTCGCACAGCGCTCGGCATCCACCAACCGCTGCCCGCCGC 2550
 L D V A Q R L G I H H R L P A P
 50 ACGCGGCCACTCCGCGCACATGGAACCCGTGGCCGCCGAGCTGCTCGCC 2600
 H A G H S A H M E P V A A E L L A
 ACCACTCGCGAGCTCCGTTACGACCGGCCCCACACCGCCATCCGAACGA 2650
 T T R E L R Y D R P H T A I P N D
 CCCCACCAACCGCCGAGTACTGGCCGAGCAGGTCCGCAACCCCGTGTGT 2700
 55 P T T A E Y W A E Q V R N P V L
 TCCACGCCACACCCAGCGTACCCGACGCCGTGGTCGAGATCGGC 2750
 F H A H T Q R Y P D A V F V E I G
 CCCGGCCAGGACCTCTCACCGCTGGTCGACGGCATGCCCTGAGAACGG 2800
 P G Q D L S P L V D G I A L Q N G
 60 CACGGCGGACGAGGTGCACGCCGCTGCACACCGCGCTGCCCGCTTCA 2850
 T A D E V H A L H T A L A R L F
 CACGCAGGCCACGCTCGACTGGTCCCGCATCCTCGGCCGGTGTGG 2900
 T R G A T L D W S R I L G G A S R
 CACGACCCCTGACGTCCCCCTCGTACCGCTCCAGCGGTCCCTACTGGAT 2950

H D P D V P S Y A F Q R R P Y W I
 CGAGTCGGCTCCCCCGGCCACGGCCGACTCGGGCCACCCCGTCCTCGGCA 3000
 E S A P P A T A D S G H P V L G
 CCGGAGTCGCCGTCGCCGGGTCGCCGGCCGGGTGTTCACGGGTCGGT 3050
 5 T G V A V A G S P G R V F T G P V
 CCCGCCGGTGCACGGCCGGTGTTCATGCCGAACTGGCGCTCGCCGC 3100
 P A G A D R A V F I A E L A L A A
 CGCCGACGCCACCGACTGCCACGGTCAACAGCTCGACGTACCTCCG 3150
 A D A T D C A T V E Q L D V T S
 10 TGCCCAGGCCGATCCGCCCGGGCAGGGCCACCGCCAGACCTGGGTCGAT 3200
 V P G G S A R G R A T A Q T W V D
 GAACCCGCCGCCGACGGCGGCCGCTTCACCGTCCACACCCGCGTCGG 3250
 E P A A D G R R F T V H T R V G
 CGACGCCCGTGGACGCTGCACGCCAGGGGGTTCTCGCCCCGGCCGCG 3300
 15 D A P W T L H A E G V L R P G R
 TGCCCCACCCCGAACGCCGTCACGCCCTGGCCCCCGCCGGCGCGGTG 3350
 V P Q P E A V D T A W P P P G A V
 CCCGCAGGGCTGCCCGGGCGTGGCGACGCCGCGGACAGGTCTTCGT 3400
 P A D G L P G A W R R A D Q V F V
 20 CGAACCGGAAGTCGACAGCCCTGACGGCTCGTGGCACACCCGACCTGC 3450
 E A E V D S P D G F V A H P D L
 TCGACGCCGTCTCTCGCGGTGACGCCGACGGGAGCCGACCCGACCGGA 3500
 L D A V F S A V G D G S R Q P T G
 25 TGGCGCACCTCGCGGTGACCGCGTCGGACGCCACCGTGTGCGCGCCTG 3550
 W R D L A V H A S D A T V L R A C
 CCTCACCCGCCGCGACAGTGGTGTGCGTGGAGCTGCCGCCTCGACGGTG 3600
 L T R R D S G V V E L A A F D G
 CCGGAATGCCGGTCTCACCGCGGAGTCGGTGACGCTGGCGAGGTGCG 3650
 30 A G M P V L T A E S V T L G E V A
 TCGGCAGGCCGATCCGACGAGTCGGACGGTCTGCTCGGCTTGAGTGGTT 3700
 S A G G S D E S D G L L R L E W L
 GCCGGTGGCGGAGGCCACTACGACGGTGCACGAGCTGCCGAGGGCT 3750
 P V A E A H Y D G A D E L P E G
 35 ACACCCCTCATCACCGCCACACACCCGACGACCCGACGACCCACCAAC 3800
 Y T L I T A T H P D D P D D P T N
 CCCCCAACACACCCACACGACCCACACACACAAACACACGCGTCCTCAC 3850
 P H N T P T R T H T Q T T R V L T
 CGCCCTCCAACACCCACCTCATCACCAACACACACCCCTCATCGTCCACA 3900
 40 A L Q H H L I T T N H T L I V H
 CCACCAACCGACCCCCCAGGCGCCGCGTACCGGCTCACCGCACCGCA 3950
 T T T D P P G A A V T G L T R T A
 CAAAACGAACACCCCGGCCATCCACCTCATCGAAACCCACACCCCA 4000
 Q N E H P G R I H L I E T H H P H
 CACCCACCTCCCCCTCACCAACTCACCAACCTCCACCAACCCACCTAC 4050
 45 T P L P L T Q L T T L H Q P H L
 GCCTCACCAACACACCCCTCACACCCCCCACCTCACCCCATCACCAAC 4100
 R L T N N T L H T P H L T P I T T
 CACCAACACACCACACAACACACCCCCACCCCTCAACCCCAA 4150
 H H N T T T T P N T P P L N P N
 50 CCACGCCATCCTCATCACCGGGCTCCGGCACCCCTGCCGCATCCTCG 4200
 H A I L I T G G S G T L A G I L
 CCCGCCACCTCAACCAACCCCCACACCTACCTCCCTCCCGCACACCA 4250
 A R H L N H P H T Y L L S R T P P
 CCCCCCACACACCCGGCACCCACATCCCTCGCACCTCACCGACCCAC 4300
 55 P P T T P G T H I P C D L T D P T
 CCAAATCACCAACGGCTCACCCACATACCACACAACCCCTCACCGGCATCT 4350
 Q I T Q A L T H I P Q P L T G I
 TCCACACCGCCGCCACCCCTCGACGACGCCACCCCTCACCAACCTCACCC 4400
 60 F H T A A T L D D A T L T N L T P
 CAACACCTCACCAACCCCCAAACCCAAAGCCGACGCCGCCCTGGCACCT 4450
 Q H L T T T L Q P K A D A A W H L
 CCACCAACACACCAACCCAAACCCACTCACCTCGTCTACTCCA 4500
 H H H T Q N Q P L T H F V L Y S
 GCGCCGCCGCCACCCCTCGGAGCCCCGGCAAGCCAACCTACGCCGCCGCC 4550

S A A A T L G S P G Q A N Y A A A
 AACGCCTCCTCGACGCCCTGCCACCCACCGCCACACCCAAGGACAACC 4600
 N A F L D A L A T H R H T Q G Q P
 CGCCACCACCATGCCCTGGGCATGTGGCACACCACCCACACTCACCA 4650
 5 A T T I A W G M W H T T T T L T
 GCCAACTCACCGACAGCGACCGCGACCGCATCCGCCGCCGGCTTCCTG 4700
 S Q L T D S D R D R I R R G G F L
 CCGATCTCGGACGACGAGGGCATGC
 P I S D D E G M

10

The *AvrII-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 12 of rapamycin is shown below.

GCATGCGGCTGTACGAGGCAGCAGGCGCACCGCGCACCGGAAGTCCC GTGGTGGT 50
 M R L Y E A A R R T G S P V V V
 15 GCGGCCGCGCTCGACGACGCCGGACGTGCCGCTGCTGCCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCGTCCGCCGTCCGGGAACGCTCTCGCCGCCACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCCGACGACGAGCGCGCACGCCCTCCCTCGCGTCG 200
 20 R S P C C P T T S A P T P P S R S
 TCCTGGAACAGCACCGCCACCGTGTGCTCGGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGACGTTCAAGGAACCTCGGCATCGACTCGCTACCGCGG 300
 P A T T T F K E L G I D S L T A
 25 TCCAGCTGCGCAACCGCGCTGACCAACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCGCGCGCGCTGCCGCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCGCGCCCGTCCGGCCGGGACCGCGGCCA 450
 30 D E L A G T R A P V A A R T A A
 CCGCGGCCGCGCACGACGAACCGCTGGCGATCGTGGGATGCCCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCGGGCGGGGTCCGGTCCGGCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 35 CGGCACCGACGCCATCACGGAGTTCCCGCGGACCGCGGCTGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCCGACGCGATCGGCAAGACCTTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCTCGACGGTGCACCGGGCTTCGACGCCGCTTCGG 700
 40 H G G F L D G A T G F D A A A F F G
 GATCAGCCCGCGCGAGGGCCCTGGGCATGGACCCCGACGAAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCCTGGGAGGGCGTTCGAAAGCGCGGGCATACCCCGACGCG 800
 L E T S W E A F E S A G I T P D A
 45 GCGCGGGCGACGCCGACACCGCGTGTTCATCGGCCGTTCTACGGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGACAGACCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGCCTCCGGCCGCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 50 S V L S G R L S Y F Y G L E G P S
 GTCAACGGTCACACCGCCCTGCTCGTCGTCACTGGTCGCCCTGCACCGAGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCTCGCGCTCGGGCGAATGCTCGCTCGCCCTGGTCGGCGGTG 1050
 G Q S L R S G E C S L A L V G G
 55 TCACGGTGATGGCGTCGCCCCGGGATTCTCGAGTTCTCCCGGACGCC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGGGGGCGAAGGGCGTTCGGCGCGGGACGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTGGTCAGCGGCTCTCCG 1200
 60 T S F A E G A G A L V V E R L S
 ACGGCGAGGCCACGGCCACACCGTCCCTCGCCCTCGTACCGGCTCCGCG 1250
 D A E R H G H T V L A L V R G S A

GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCCGAACGGCCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCGTCATCCACCAAGGCCCTCGCGAACCGCGAAACTCACCCCCG 1350
 5 Q E R V I H Q A L A N A K L T P
 CCGATGTGGACCGGGTCGAGGCCACGGCACCCGCCCTGGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCAGGGCGCTGCTCGCGACGTACGGACAGGACGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 GCCCCCTGCTGCTCGGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCCG 1500
 10 P L L L G S L K S N I G H A Q A
 CGTCAGGGGGTCGCCGGGATCATCAAGATGGTCGAGGCCATCGGCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGCCGAGGCCGTCGCCGACGTCGACTG 1600
 E L P P T L H A D E P S P H V D W
 15 GACGGCCGGTGCCTGAGCTCTGACGTCCGCCGGCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTCGCCCTAGGCCGGCAGGCCGTGTCGTCCTCGGGATCAGTGGCACC 1700
 T G R P R R A G V S S F G I S G T
 AACGCCACGTCATCTGGAAAGCGCACCCCCACTCAGCTGGACAA 1750
 20 N A H V I L E S A P P T Q P A D N
 CGCGGTGATCGAGCCGGCACCGGAGTGGGTGCCGTTGGTGAATTCCGGCA 1800
 A V I E R A P E W V P L V I S A
 GGACCCAGTCGGCTTGACTGAGCACGCCGGTTGCGTGCATCTG 1850
 R T Q S A L T E H E G R L R A Y L
 25 GCGGCCTGCCCGGGGTGGATATGCCGGCTGTCGACGCTGGCGAT 1900
 A A S P G V D M R A V A S T L A M
 GACACGGTCGGTGTTCGAGCACCGTGCCGTCGCTGGAGATGACACCG 1950
 T R S V F E H R A V L L G D D T
 TCACCGGCACCCCTGTGTCGACCCCTGGCGGTGTTCGTCTTCCCGGG 2000
 30 V T G T A V S D P R A V F V F P G
 CAGGGGTCCAGCGTCTGGCATGGGTGAGGAACCTGGCCGGCTTCCC 2050
 Q G S Q R A G M G E E L A A A A F P
 CGTCTCGCCGGATCCATCAGCAGGTGTGGACCTGCTCGATGTGGCCG 2100
 V F A R I H Q Q V W D L L D V P
 35 ATCTGGAGGTGAACGAGACCGTTACGCCACGCCCTGTTCGCAATG 2150
 D L E V N E T G Y A Q P A L F A M
 CAGGTGGCTCTGTTGGCTGCTGGAAATCGGGGTGACGACCGGACGC 2200
 Q V A L F G L L E S W G V R P D A
 GGTGATCGGCCATTGGTGGGTGAGCTTGCGCTGCGTATGTGTCGGGG 2250
 40 V I G H S V G E L A A A Y V S G
 TGTGGTCGGAGGATGCCCTGCACTTTGGTGTGGCGCGGGCTCGTCTG 2300
 V W S L E D A C T L V S A R A R L
 ATGCAGGCTCTGCCCGGGTGGGTGATGGTCGCTGTCGGCTCGGA 2350
 M Q A L P A G G V M V A V P V S E
 45 GGATGAGGCCGGCGCTGCTGGGTGAGGTGAGATGCCCGGGTCA 2400
 D E A R A V L G E G V E I A A V
 ACGGCCGTCGTCGGTGGTTCTCTCCGGTGTGGAGGCGCCGTGCTGCAG 2450
 N G P S S V V L S G D E A A V L Q
 GCCCGGGAGGGCTGGGAAGTGGACGCCGGTGGCGACCGCACCGTT 2500
 50 A A E G L G K W T R L A T S H A F
 CCATTCCGCCGTATGGAACCCATGCTGGAGGAGTTCCGGCGGGTCA 2550
 H S A R M E P M L E E F R A V A
 AAGGCCTGACCTACCGGACGCCGAGGTCTCCATGGCGTTGGTGTACAG 2600
 E G L T Y R T P Q V S M A V G D Q
 55 GTGACCAACCGCTGAGTACTGGGTGCGGCCAGGTCCGGACACGGTCCGGT 2650
 V T T A E Y W V R Q V R D T V R F
 CGGCGAGCAGGTGGCCTCGTACGAGGACGCCGTGTCGAGCTGGGT 2700
 G E Q V A S Y E D A V F V E L G
 CCGACCGGTCACTGGCCCGCTGGTCGACGGTGTGCGATGCTGCCACGGC 2750
 60 A D R S L A R L V D G V A M L H G
 GACCACGAAATCCAGGCCGCGATCGGCCCTGGCCCACCTGTATGTCAA 2800
 D H E I Q A A I G A L A H L Y V N
 CGCGTCACGGTCGACTGGCCCGCTCTGGCGATGCTCCGGCAACAC 2850
 G V T V D W P A L L G D A P A T

GGGTGCTGGACCTTCCGACATACGCCCTCAGCACCAAGCGCTACTGGCTC 2900
 R V L D L P T Y A F Q H Q R Y W L
 GAGTCGGCTCCCCCGGCCACGGCGACTCGGGCCACCCCGTCTCGGCAC 2950
 E S A P P A T A D S G H P V L G T
 5 CCGAGTCGCCGTGCCGGGTCGCCGGGCGGGTGTTCACGGGTCCCGTGC 3000
 G V A V A G S P G R V F T G P V
 CCGCCGGTGGGACCGCGCGGTGTTCATGCCGAACTGGCGCTCGCCGCC 3050
 P A G A D R A V F I A E L A L A
 GCGCACGCCACCGACTGCCACGGTCGAACAGCTCGACGTACCTCCGT 3100
 10 A D A T D C A T V E Q L D V T S V
 GCCCCGGCGATCCGCCCGGGCAGGGCACCGCCAGACCTGGTGTGATG 3150
 P G C S A R G R A T A Q T W V D
 AACCCGCCCGACGGCGCGCTTCACCGTCCACCCCGCGTCGGC 3200
 15 E P A A D G R R F T V H T R V G
 GACGCCCGTGGACGCTGCACGCCGAGGGGTTCTCCGCCCCGGCGCGT 3250
 D A P W T L H A E G V L R P G R V
 GCGCCAGCCGAAGCCGTCGACACCGCTGGCCCCGGCGCGGTGC 3300
 P Q P E A V D T A W P P P G A V
 CCGCGGACGGGCTGCCGGGCGTGGCGACGCCGCGGACAGGTCTCGTC 3350
 20 P A D G L P G A W R R A D Q V F V
 GAAGCCGAAGTCGACAGCCCTGACGGCTTCGTGGCACACCCGACCTGCT 3400
 E A E V D S P D G F V A H P D L L
 CGACCGGGCTTCTCCCGGGTCCGGCGACGGGAGCCGCCAGCCGACCGGAT 3450
 D A V F S A V G D G S R Q P T G
 25 GGCACCGACCTCGCGGTGACCGCTGGACGCCACCGTGTGCGCGCCTGC 3500
 W R D L A V H A S D A T V L R A C
 CTCACCCGCCGCGACAGTGGTGTGAGCTCGCCGCTTCGACGGTGC 3550
 L T R R D S G V V E L A A F D G A
 CGGAATGCCGGTGTCTACCGCGGAGTCGGTGACGCTGGCGAGGTGCGGT 3600
 30 G M P V L T A E S V T L G E V A
 CGGCAGGCCGATCCGACGAGTCGGACGGTGTGCTTCGGTTGAGTGGTTG 3650
 S A G G S D E S D G L L R L E W L
 CCGGTGGCGGAGGCCACTACGACGGTGCCGACGAGCTGCCGAGGGCTA 3700
 P V A E A H Y D G A D E L P E G Y
 35 CACCCCTCATCACGCCACACACCCCGACGACCCCCGACGACCCCCACCAACC 3750
 T L I T A T H P D D P D D P T N
 CCCACAAACACACCCACACGACCCACACACAAACACACCGTCTCACC 3800
 P H N T P T R T H T Q T R V L T
 40 GCCCTCAACACCCACCTCATCACACCAACACACCCCTCATCGTCCACAC 3850
 A L Q H H L I T T N H T L I V H T
 CACCAACGCCGACCCCGGCGCCGTCACCGGCTCACCCGACCGCAC 3900
 T T D P P G A A V T G L T R T A
 AAAACGAACACCCCGGCCATCCACCTCATCGAAACCCACACCCCCAC 3950
 Q N E H P G R I H L I E T H H P H
 45 ACCCCACTCCCCCTCACCAACTCACCAACCCCTCACCAACCCCCACCTACG 4000
 T P L P L T Q L T T L H Q P H L R
 CCTCACCAACACCCCTCCACACCCCCCACCTCACCCCCATCACCAACCC 4050
 L T N N T L H T P H L T P I T T
 ACCACAAACACCAACCAACCCACACCCCCAACACCCCCACCCCTCAACCCCAAC 4100
 50 H H N T T T T P N T P P L N P N
 CACGCCATCCTCATCACGGCGGCTCCGGCACCCCTCGCCGGCATCCTCGC 4150
 H A I L I T G G S G T L A G I L A
 CCGCCACCTCAACCCACCCCCACACCTACCTCTCTCCCGCACACCAACCC 4200
 R H L N H P H T Y L L S R T P P
 55 CCCCCACCAACCCGGCACCCACATCCCCCTGCGACCTCACCGACCCCCACC 4250
 P P T T P G T H I P C D L T D P T
 CAAATCACCAAGCCCTACCCACATACCCACAACCCCTCACCGGCATCTT 4300
 Q I T Q A L T H I P Q P L T G I F
 CCACACCGCCGCCACCCCTCGACGACGCCACCCCTCACCAACCTCACCCCCC 4350
 60 H T A A T L D D A T L T N L T P
 AACACCTCACCAACCCCTCAACCCAAAGCCGACGCCCTGGCACCTC 4400
 Q H L T T T L Q P K A D A A W H L
 CACCAACACCCAAAACCAACCCCTCACCCACTTCGCTCTACTCCAG 4450
 H H H T Q N Q P L T H F V L Y S S

CGCCGCCGCCACCCCTGGCAGCCCCGGCCAAGCCAACCTACGCCGCC 4500
 A A A T L G S P G Q A N Y A A A
 ACGCCTTCCTCGACGCCCTGCCACCCACCGCCACACCCAAGGACAACCC 4550
 N A F L D A L A T H R H T Q G Q P
 5 GCCACCACCATGCCCTGGGCATGTGGCACACCACCAACTCACCAG 4600
 A T T I A W G M W H T T T L T S
 CCAACTCACCGACAGCGACCGGACCGCATTCCGCCGGCGCTTCCTGC 4650
 Q L T D S D R D R I R R G G F L
 CGATCTCGACGACGAGGGCATGC
 10 P I S D D E G M

The *AvrII-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 13 of rapamycin is shown below.

GCATGCGGTGTACGAGGCACGGCACCGGAAGTCCGTGGTGGT 50
 15 M R L Y E A A R R T G S P V V V
 GCGGCCGCGCTCGACGACGCGCCGGACGTGCCGTGCGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCCTCGCGTGCCTGGGAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 20 GCTCGCCGTGCTGCCGACGAGCGCGGCCACGCCCTCGCGTTG 200
 R S P C C P T T S A P T P P S R S
 TCCTGGAACAGCACCGCCACCGTGCTCGGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACTCGGCATCGACTCGTCACCGCGG 300
 25 P A T T T F K E L G I D S L T A
 TCCAGCTGCGAACCGCGTACCGACGGCGACCGGGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCGCGCGCTCGCCGCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 30 CGACGAGCTGCCGGTACCCGCCGCCGCGCCCGGACCGCGGCCA 450
 D E L A G T R A P V A A R T A A
 CCGCGCCGCGCACGACGAACCGCTGGCGATCGTGGCATGCCCTGCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGCGGGTCCGCTCGCCACAGGAGCTGTGGCGTCGCTCGTC 550
 35 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATACGGAGTTCCCGCGGACCGCGGCTGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTTACGACCCGGACCCCGACGCGATCGCAAGACCTCGTCCGG 650
 40 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCCTCGACGGTGCACCGGCTTCGACGCCGGCTTCGG 700
 H G G F L D G A T G F D A A A F F G
 GATCAGCCCGCGAGGCCCTGGCATGGACCCGACGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCCGGAGGCCTCGAAAGCCGGCATCACCCGGACGCG 800
 45 L E T S W E A F E S A G I T P D A
 GCGCGGGCAGCGACACCGCGTGTTCATCGCGCGTCTCTACGGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGACAGACCA 900
 G T G A D T N G F G A T G S Q T
 50 GCGTGTCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 GTCACGGTGCACACCCCTGCTCGTCGTCAGTGGTCGCCCTGCACCGAGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGTAGCTCCCTGCCCTGGCGGAATGCTCGCTGCCCTGGTGGCGGTG 1050
 55 G Q S L R S G E C S L A L V G G
 TCACGGTGTGGCGTCGCCCGGGATTCGTCGAGTTCTCCGGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGGGCGGGCGAAGGGCTTCGGCGGGCGGACGG 1150
 G L A P D G R A K A F G A G A D G
 60 TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTGGTCGAGCGGCTCTCCG 1200
 T S F A E G A G A L V V E R L S
 ACGGCGAGCGCCACGGCCACACCGTCCCTCGCCCTCGTACGCGGCTCCGCG 1250

D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCCGAACGGCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCGTCATCCACCAAGGCCCTCGCGAACGCCAAACTCACCCCCG 1350
 5 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACCGCGTCAGGCCACGGCACCGGACCCGCTCGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCAGGCCGCTGCTCGCACGGTACGGACAGGACCGGGCAG 1450
 10 P I E A Q A L L A T Y G Q D R A T
 GCCCCCTGCTGCTCGGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCC 1500
 P L L L G S L K S N I G H A Q A
 CGTCAGGGGTCGCCGGGATCATCAAGATGGTGCAGGCCATCGGCCACGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGCCAGGCCGTGCCGCACGTCGACTG 1600
 15 E L P P T L H A D E P S P H V D W
 GACGGCCGGTGCCGTCGAGCTCTGACGTCGGCCGGCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTCGCCCTAGGCCGGCGGGCGTGTGTCGCTCTGGAGTCAGGCCACC 1700
 T G R P R R A G V S S F G V S G T
 20 AACGCCAACGICATCTGGAGAGCCGACCCCCCGCTCAGCCCGGGAGGA 1750
 N A H V I L E S A P P A Q P A E E
 GGCGCAGCTGTTGAGACGCCGGTGGTGGCTCGGATGTGCTGCCGCTGG 1800
 A Q P V E T P V V A S D V L P L
 TGATATCGGCCAAGACCCAGGCCGCCCTGACCGAACACGAAGACGGCTG 1850
 25 V I S A K T Q P A L T E H E D R L
 CGCGCCTACCTGGCGGCCGTCGCCGGGCGGATATACGGCTGTCGGCATC 1900
 R A Y L A A S P G A D I R A V A S
 GACGCTGGCGGTGACACGGTCGGTGGTCGAGCACCGGCCGTACTCCTG 1950
 T L A V T R S V F E H R A V L L
 30 GAGATGACACCGTCACCGCACCGCGGTGACCGACCCAGGATCGTGT 2000
 G D D T V T G T A V T D P R I V F
 GTCTTCCCAGGGCAGGGGTGGCAGTGGCTGGGATGGCAGTGCACTGCG 2050
 V F P G Q G W Q W L G M G S A L R
 CGATTGTCGGTGGTGGTCGCCGAGCGGATGGCCGAGTGTGCGGCCGT 2100
 35 D S S V V F A E R M A E C A A A
 TGCGCAGTCTGGACTGGGATCTGTTACGGGTTCTGGATGATCCGGCG 2150
 L R E F V D W D L F T V L D D P A
 GTGGTGGACCGGGTTGATGTGGTCCAGGCCGCTCTGGCGATGATGGT 2200
 V V D R V D V V Q P A S W A M M V
 40 TTCCCTGGCCGCGGTGGCAGGCCGGTGTGCGGCCGGATGCGGTGA 2250
 S L A A V W Q A A G V R P D A V
 TCGGCCATTGCAAGGGTGAGATCGCCGAGCTTGTGTCGGCGGGTGC 2300
 I G H S Q G E I A A A C V A G A V
 TCACTACGCGATGCCGCCGGATCGTGAACCTTGCAGCCAGGGATCGC 2350
 45 S L R D A A R I V T L R S Q A I A
 CCGGGGCTGGCGGGCGGGCGCGATGGCATCCGTGCCCTGCCCGCGC 2400
 R G L A G R G A M A S V A L P A
 AGGATGTCGAGCTGGTCGACGGGCGTGGATCGCCGCCACACGGGCC 2450
 Q D V E L V D G A W I A A H N G P
 50 GCCTCCACCGTGATCGCGGGCACCCCGGAAGCGTCGACCATGTCCTC 2500
 A S T V I A G T P E A V D H V L T
 CGCTCATGAGGCACAAGGGGTGCGGGTGCAGGCCGATACCGTCGACTATG 2550
 A H E A Q G V R V R R I T V D Y
 CCTCGCACACCCCGCACGTCGAGCTGATCCCGCACGAACACTCGACATC 2600
 55 A S H T P H V E L I R D E L L D I
 ACTAGCGACAGCAGCTCGCAGACCCCGCTCGTGCCTGGCTGTCGACCGT 2650
 T S D S S S Q T P L V P W L S T V
 GGACGGCACCTGGGTGACAGCCCGCTGGACGGGGAGTACTGGTACCGGA 2700
 D G T W V D S P L D G E Y W Y R
 60 ACCTGCGTGAACCGGTCGGTTCCACCCGCCGTCAAGCCAGTTGCAGGCC 2750
 N L R E P V G F H P A V S Q L Q A
 CAGGGCGACACCGTGGTCGAGGTCAAGGCCAGCCGGTGGTGTGCA 2800
 Q G D T V F V E V S A S P V L L Q
 GGCGATGGACGACGATGTCGTACGGTTGCCACGCTGCGTCGTGACGACG 2850

A M D D D V V T V A T L R R D D
 GCGACGCCACCCGGATGCTCACCGCCCTGGCACAGGCCTATGTCCACGGC 2900
 G D A T R M L T A L A Q A Y V H G
 GTCACCGTCGACTGGCCGCCATCCTCGGCACCACCAACCCGGGTACT 2950
 5 V T V D W P A I L G T T T T R V L
 GGACCTTCCGACCTACGGCTTCCAACACCAGCGGTACTGGCTCGAGTCGG 3000
 D L P T Y A F Q H Q R Y W L E S
 CTCCCCCGGCCACGGCCGACTCGGGCCACCCCGTCTCGGCACCGGAGTC 3050
 10 A P P A T A D S G H P V L G T G V
 GCCGTGCGGGGGTGCACGGGGGTGTTCACGGGTCCCCTGCCCCGGG 3100
 A V A G S P G R V F T G P V P A G
 TGCGGACCGCGCGGTGTTCATCGCGAACCTGGCGCTCGCCGCCGACG 3150
 A D R A V F I A E L A L A A A D
 15 CCACCGACTGCGCCACGGTCGAACAGCTCGACGTACCTCCGTGGCCGG 3200
 A T D C A T V E Q L D V T S V P G
 GGATCCGCCCGGGCAGGGCCACCGCGCAGACCTGGTCGATGAACCCGC 3250
 G S A R G R A T A Q T W V D E P A
 CGCCGACGGGCGGCCGCTCACCGTCCACACCCCGCTGGCGACGCC 3300
 A D G R R R F T V H T R V G D A
 20 CGTGGACGCTGCACGCCGAGGGGTTCTCCGCCCCGGCGTGGCCAG 3350
 P W T L H A E G V L R P G R V P Q
 CCCGAAGCCGTCGACACCGCCTGGCCCCCGCCGGCGCGGTGCCCCCGGA 3400
 P E A V D T A W P P P G A V P A D
 CGGGCTGCCGGGGCGTGGCGACCGCGCGGACCAGGTCTCGTGAAGCCG 3450
 25 G L P G A W R R A D Q V F V E A
 AAGTCGACAGCCCTGACGGCTTCGTGGCACACCCCGACCTGCTCGACGCG 3500
 E V D S P D G F V A H P D L L D A
 GTCTTCTCCGCGGTGCGACGGGAGCCGACCGGATGGCGCGA 3550
 V F S A V G D G S R Q P T G W R D
 30 CCTCGCGGTGCACCGTGGACGCCACCGTGTGCGCGCCCTGCCTCACCC 3600
 L A V H A S D A T V L R A C L T
 GCCCGCACAGTGGTGTGCGAGCTCGCCGCTTCGACGGTGCCGGAATG 3650
 R R D S G V V E L A A F D G A G M
 CGGGTGTCAACCGCGGAGTCGGTACGCTGGCGAGGTGCGTCGGCAGG 3700
 35 P V L T A E S V T L G E V A S A G
 CGGATCCGACGAGTCGGACGGTCTGCTGGCTTGAGTGGTTGCCGGTGG 3750
 G S D E S D G L L R L E W L P V
 CGGAGGCCCACTACGACGGTGCCGAGCTGCCGAGGGCTACACCCCTC 3800
 A E A H Y D G A D E L P E G Y T L
 40 ATCACCGCCACACACCCGACGACCCGACGACCCACCAACCCACAA 3850
 I T A T H P D D P D D P T N P H N
 CACACCCACACGCAACACACACAAACACACCGTCCCTCACCGCCCTCC 3900
 T P T R T H T Q T T R V L T A L
 AACACCACTCATCACCAACACCCCTCATCGTCCACACCAACCCACC 3950
 45 Q H H L I T T N H T L I V H T T T
 GACCCCCCAGGCGCCGTCACCGGCTCACCCGACCGCACAAACCGA 4000
 D P P G A A V T G L T R T A Q N E
 ACACCCCGGCCGATCCACCTCATCGAAACCCACACCCACACCCAC 4050
 H P G R I H L I E T H H P H T P
 50 TCCCCCTCACCCAACTCACCAACCCCTCCACCAACCCACCTACGCCCTACC 4100
 L P L T Q L T T L H Q P H L R L T
 AACAAACACCCCTCCACACCCCCCACCTCACCCCATCACCAACCCACAA 4150
 N N T L H T P H L T P I T T H H N
 CACCAACCACAAACCAACCCCCAACACCCACCCCTCAACCCCAACCGCCA 4200
 55 T T T T T P N T P P L N P N H A
 TCCTCATCACCGCGGCTCCGGCACCCCTCGCCGGCATCTCGCCGCCAC 4250
 I L I T G G S G T L A G I L A R H
 CTCAACCACCCCCACACCTACCTCTCTCCCGACACCACCAACCCAC 4300
 L N H P H T Y L L S R T P P P P T
 60 CACACCCGGCACCCACATCCCTGCGACCTCACCGACCCACCCAAATCA 4350
 T P G T H I P C D L T D P T Q I
 CCCAAGCCCTCACCCACATACCGACCAACCCCTCACCGGATCTTCCACACC 4400
 T Q A L T H I P Q P L T G I F H T
 GCCGCCACCCCTGACGACGCCACCCCTCACCAACCTCACCCCCAACACCT 4450

```

A A T L D D A T L T N L T P Q H L
CAACCACCCCTCCAACCAAAGCCGACGCCGCTGGCACCTCCACCACC 4500
T T T L Q P K A D A A W H L H H
5 ACACCCAAAACCAACCCCTACCCACTTCGTCCTCTACTCCAGCGCCGCC 4550
H T Q N Q P L T H F V L Y S S A A
GCCACCCCTCGGAGCCCCGGCCAAGCCAACCTACGCCGCCAACGCCCT 4600
A T L G S P G Q A N Y A A A A N A F
CCTCGACGCCCTGCCACCCACCGCCACACCCAAGGACAACCCGCCACCA 4600
L D A L A T H R H T Q G Q P A T
10 CCATCGCCTGGGCATGTGGCACACCACCACTCACCAGCCAACTC 4700
T I A W G M W H T T T L T S Q L
ACCGACAGCGACCGCGACCGCATCCGCCGCCGGCTTCCTGCCATCTC 4750
T D S D R D R I R R G G F L P I S
GCACGACGAGGGCATGC
15 D D E G M

```

The *NheI-XhoI* hybrid FK-506 PKS module 8 containing the AT domain of module 12 of rapamycin is shown below.

```

GCATGCGGCTGTACGAGGGCGCACGGCGCACCGGAAGTCCGTGGTGGT 50
20 M R L Y E A A R R T G S P V V V
GCGGCCGCGCTCGACCGACGCCGGACGTGCCGCTGCGCGGGCTGCG 100
A A A L D D A P D V P L L R G L R
GCGTACGACCGTCCGGCGTCCGGCGTCCGGGAACGCTCTCGCCGACC 150
R T T V R R A A V R E R S L A D
25 GCTCGCCGTGCTGCCGACGACGAGCGCGGCCACGCCCTCGCGTTCG 200
R S P C C P T T S A P T P P S R S
TCCTGGAACAGCACCGCCACCGTGTGCGGCCACCTGGCGCCAGACAT 250
S W N S T A T V L G H L G A E D I
CCC GGCGACGACGACGTTCAAGGAACCTGGCATCGACTCGCTACCGCGG 300
30 P A T T T F K E L G I D S L T A
TCCAGCTGCGCAACGCGCTGACCAACGGCGACCGGGCTACGCCAACGCC 350
V Q L P N A L T T A T G V R L N A
ACAGCGGTCTTCGACTTCCGACGCCGCCGCTGCCCGAGACTCGG 400
T A V F D F P T P R A L A A R L G
35 CGACGAGCTGGCCGGTACCCGCGGCCGTCGCGGGCCCGACCGGGCCA 450
D E L A G T R A P V A A R T A A
CCGCAGCCGCGCACGACGAACCGCTGGCGATCGTGGGCATGGCCTGCCGT 500
T A A A H D E P L A I V G M A C R
40 CTGCCGGCGGGGTGCGTCGCCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
L P G G V A S P Q E L W R L V A S
CGGCACCGACGCCATCACGGAGTTCCCCGGACCGGGCTGGACGTGG 600
G T D A I T E F P A D R G W D V
ACCGCGCTCTACGACCCGGACCCCGACGCGATCGGCAAGACCTTCGTCGG 650
D A L Y D P D P D A I G K T F V R
45 CACGGCGGCTTCCTCGACGGTGCACCGGCTTCGACGCCGGCTTCGG 700
H G G F L D G A T G F D A A A F F G
GATCAGCCCCCGCGAGGCCCTGGCCATGGACCCCGACCAACGGGTGCTCC 750
I S P R E A L A M D P Q Q R V L
TGGAGACGTCCTGGGAGGGCTTCGAAAGCGCGGGCATACCCCGACGCC 800
50 L E T S W E A F E S A G I T P D A
GCGCGGGCGAGCGACACCGGCTGTTCATCGGCCGCTTCCTACGGGTA 850
A R G S D T G V F I G A F S Y G Y
CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGAGACCA 900
G T G A D T N G F G A T G S Q T
55 GCGTGCTCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
S V L S G R L S Y F Y G L E G P S
GTCACGGTCGACACCGCCTGCTCGTCGTCACTGGTCGCCCTGCACCCAGGC 1000
V T V D T A C S S S L V A L H Q A
AGGGCAGTCCCTCGCGCTGGGCGAATGCTCGCTGCCCTGGTCGGCGGTG 1050
60 G Q S L R S G E C S L A L V G G
TCACGGTGATGGCGTCGCCGGGAGTCGAGTTCTCCCGCAGCGC 1100
V T V M A S P G G F V E F S R Q R

```

GGGCTCGCGCCGGACGGGCGGGCGAAGGCCTTCGGCGCGGGCGGGACGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCAGGGCGCCGGTGCCTGGTGGTCGAGCGGGCTCTCCG 1200
 5 T S F A E G A G A L V V E R L S
 ACGCGGAGCGCCACGGCACACCGTCTCGCCCTCGTACCGGGCTCCGCG 1250
 D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCTGAACGGTCTGCGCGCCGAACGGCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCTCATCCACCAGGCCCTCGGAACCGAAACTCACCCCCG 1350
 10 Q E R V I H Q A L A N A K L T P
 CCGATGTGCGACGCCGGTCGAGGCCACGGCACCCGCCCTGGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCAGGCCGCTCGCGACGTACGGACAGGACGGGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 15 GCCCCCTGCTGCCTCGCTGAAGTCGAACATCGGGCACGCCAGGCCG 1500
 P L L L G S L K S N I G H A Q A
 CGTCAGGGGTGCGCCGGATCATCAAGATGGTGCAGGCCATCCGGCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGCCGACGAGCCGTCGCCGCACGTCGACTG 1600
 20 E I P P T L H A D E P S P H V D W
 GACGGCCGGTGCCGCTCGAGCTCTGACGTCGGCCGGCCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGTGCCTCCGCCGCGCCGCTGCCGTCTCGTCTCGTGGCGTGGCGGACG 1700
 T G R P R R A A V S S F G V S G T
 25 AACGCCACATCATCCTGAGGCAGGACCGGTCAAAACGGGACCGGTCA 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCGATCGAGGCAGGACCGGTCAAGTAGGACCGGTGAGGCTG 1800
 A G A I E A G P V E V G P V E A
 GACCGCTCCCCGCCGGCGCCGCCGTACGACCGGGCGAAGACCTTCCGCTG 1850
 30 G P L P A A P P S A P G E D L P L
 CTCGTGCGCCGCTTCCCCGGAGGCACTCGACGAGCAGATCGGGCGCT 1900
 L V S A R S P E A L D E Q I G R L
 GCGCGCCTATCTGACACCGGCCGGCGTCGACCGGGCGCCGTGGCGC 1950
 R A Y L D T G P G V D R A A V A
 35 AGACACTGGCCGGCGTACGCACCTCACCCACCGGGCGTACTGCTCGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTACCGCCGCTCCCCCGCGGACCGAGCCGACGAACTCGTCTT 2050
 D T V I G A P P A D Q A D E L V F
 CGTCTACTCCGGTCAGGGCACCCAGCATCCGCGATGGCGAGCAGCTAG 2100
 40 V Y S G Q G T Q H P A M G E Q L
 CCGCCGCGTTCCCCGTCTCGCGCGATCCATCAGCAGGTGTGGACCTG 2150
 A A A F P V F A R I H Q Q V W D L
 CTCGATGTGCCCAGTCGGAGGTGAACGAGACCGGTACGCCAGCCGGC 2200
 L D V P D L E V N E T G Y A Q P A
 45 CCTGTTCGCAATGCAGGTGGCTCTGTTGGCTGCTGGAACTGTTGGG 2250
 L F A M Q V A L F G L L E S W G
 TACGACCGGACGCCGGTGAATCGGCCATTGCGTGGGTGAGCTGGCGCTG 2300
 V R P D A V I G H S V G E L A A A
 TATGTGTCCGGGTGTGGCTGGAGGATGCCGACTTGGTGTGGCG 2350
 50 Y V S G V W S L E D A C T L V S A
 GCGGGCTCGTCTGATGCAGGCTCTGCCGCCGGTGGGTGATGGTCGCTG 2400
 R A R L M Q A L P A G G V M V A
 TCCCGGTCTCGGAGGATGAGGCCGGCGTGGTGGGTGAGGGTGTGGAG 2450
 V P V S E D E A R A V L G E G V E
 55 ATCGCCGCGGTCAACGCCCGTCGCTGGTGGTCTCTCCGGTGAAGGC 2500
 I A A V N G P S S V V L S G D E A
 CGCCGTGCTGCAGGCCGGAGGGCTGGGAAGTGGACGCCGGCTGGCGA 2550
 A V L Q A A E G L G K W T R L A
 CCAGCCACCGCTTCCATTCCGCCGTATGGAACCATGCTGGAGGAGTTC 2600
 60 T S H A F H S A R M E P M L E E F
 CGGGCAGGTGCGCCGAAGGCCTGACCTACCGGACGCCGAGGTCTCCATGGC 2650
 R A V A E G L T Y R T P Q V S M A
 CGTTGGTGAATCAGGTGACCACCGCTGAGTACTGGGTGCGGCAGGTCCGGG 2700
 V G D Q V T T A E Y W V R Q V R

ACACGGTCCGGTTGGCGAGCAGGTGGCCTCGTACGAGGACGCCGTGTC 2750
 D T V R F G E Q V A S Y E D A V F
 GTCGAGCTGGGTGCCGACCGGTCACTGGCCGCCCTGGTCGACGGTGTGCGC 2800
 V E L G A D R S L A R L V D G V A
 5 GATGCTGCACGGCGACCAACGAAATCCAGGCCGATCGGCCGCCCTGGCCC 2850
 M L H G D H E I Q A A I G A L A
 ACCTGTATGTCACCGCGTACGGTCACTGGCCGCCCTGGCGAT 2900
 H L Y V N G V T V D W P A L L G D
 GCTCCGGCAACACGGGTGCTGGACCTTCCGACATACGCCCTCCAGCACCA 2950
 10 A P A T R V L D L P T Y A F Q H Q
 GCGCTACTGGCTCGAGTCGGCTCCCCCGGCCACGGCCGACTCGGGCCACC 3000
 R Y W L E S A P P A T A D S G H
 CCGTCCTCGGCACCGGAGTCGCCGTGCCGGTCGCCGGCGGGTGTTC 3050
 P V L G T G V A V A G S P G R V F
 15 ACGGGTCCCGTGCCGCCGGTGCGGACCGCGCGGTGTTCATGCCGA 3100
 T G P V P A G A D R A V F I A E L
 GCGCCTGCCGCCGCCGACGCCACCGACTGCCACGGTCAACAGCTCG 3150
 A L A A A D A T D C A T V E Q L
 ACGTCACCTCCGTGCCCGCGGATCCGCCGCCGGCAGGGCACCGCGCAG 3200
 20 D V T S V P G G S A R G R A T A Q
 ACCTGGGTGATGAACCCGCCGACGGGGCGGCCCTCACCCTCA 3250
 T W V D E P A A D G R R R F T V H
 CACCCCGCTGGCGACGCCGGTGACGCTGACGCCAGGGGTTCTCC 3300
 T R V G D A P W T L H A E G V L
 25 GCCCCGGCCGCGTGCCCGAGCCCGAAGCCGTCGACACCCCTGGCCCCCG 3350
 R P G R V P Q P E A V D T A W P P
 CCGGGCGCGGTGCCGCCGGACGGGCTGCCGGCGTGGCGACGCCGGA 3400
 P G A V P A D G L P G A W R R A D
 CCAGGTCTTCGTCGAAGCCGAAGTCGACAGCCCTGACGGCTTCGTGGCAC 3450
 30 Q V F V E A E V D S P D G F V A
 ACCCCGACCTGCTCGACGCCGGTCTCTCCGCCGACGGGAGCCGC 3500
 H P D L L D A V F S A V G D G S R
 CAGCCGACCGGATGGCGCGACCTCGCCGGTGCACGCCGCGACGCCACCGT 3550
 Q P T G W R D L A V H A S D A T V
 35 GCTGCGCGCCTGCCCTACCCGCCGACAGGGTGTGGAGCTCGCCG 3600
 L R A C L T R R D S G V V E L A
 CCTTCGACGGTGCCGGAATGCCGGTGCACCGCGGAGTCGGTGACGCTG 3650
 A F D G A G M P V L T A E S V T L
 GGCAGGGTCCGCGTCCGCCGAGATCCGACGGAGTCGGACGGTCTGCTTC 3700
 40 G E V A S A G G S D E S D G L L R
 GCTTGAGTGGTTGCCGGTGGCGGAGGCCACTACGACGGTGCCGACGAGC 3750
 L E W L P V A E A H Y D G A D E
 TGCCCGAGGGCTACACCCCTCATCACCGCCACACACCCGACGCCGAC 3800
 L P E G Y T L I T A T H P D D P D
 45 GACCCCACCAACCCCCACAACACACCCACACGCCACACACAAACAC 3850
 D P T N P H N T P T R T H T Q T T
 ACGCGTCCCTCACCGCCCTCCAACACCCACCTCATCACCAACACACCC 3900
 R V L T A L Q H H L I T T N H T
 TCATCGTCCACACCACCGACCCCCCAGGCCGCCGTCACCGGCC 3950
 50 L I V H T T T D P P G A A V T G L
 ACCCGCACCGCACAAACGAACACCCGGCCGATCCACCTCATCGAAC 4000
 T R T A Q N E H P G R I H L I E T
 CCACCAACCCCCACACCCACTCCCCCTCACCAACTCACCAACCCCTCC 4050
 H H P H T P L P L T Q L T T L H
 55 AACCCCACCTACGCCCTACCAACAACACCCCTCACACCCCCCACCTCACC 4100
 Q P H L R L T N N T L H T P H L T
 CCCATCACCAACCCACCAACACCCACACCAACCCCCAACACCCACC 4150
 P I T T H H N T T T T P N T P P
 CCTCAACCCCAACCACGCCATCCATCACCGGCCGGCTCCGGCACCCCTCG 4200
 60 L N P N H A I L I T G G S G T L
 CCGGCATCCTCGCCGCCACCTCAACCCACACCCACACCTACCTCCTCTCC 4250
 A G I L A R H L N H P H T Y L L S
 CGCACACCAACCCCCACCAACACCCGGCACCCACATCCCCTGCGACCT 4300
 R T P P P P T T P G T H I P C D L

CACCGACCCCACCCAAATCACCAAGCCCTCACCCACATACCACAACCCC 4350
 T D P T Q I T Q A L T H I P Q P
 TCACUGGCATCTTCCACACCGCCGCCACCCCTCGACGACGCCACCCCTCACC 4400
 L T G I F H T A A T L D D A T L T
 5 AACCTCACCCCCCAACACCTCACCCACCACCCAAAGCCGACGC 4450
 N L T P Q H L T T T L Q P K A D A
 CGCCTGGCACCTCCACCACCAACCCAAACCCCTCACCCACTTCG 4500
 A W H L H H T Q N Q P L T H F
 TCCTCTACTCCAGCGCCGCCACCCCTCGGAGCCGGCAAGCCAAC 4550
 10 V L Y S S A A A T L G S P G Q A N
 TAGCCGCCAACGCCCTTCCTCGACGCCCTGCCACCCACGCCACAC 4600
 Y A A A N A F L D A L A T H R H T
 CCAAGGACAACCCGCCACCACCATGCCCTGGGCATGTGGCACACCA 4650
 Q G Q P A T T I A W G M W H T T
 15 CCACACTCACCAAGCCAACTCACCGACAGCGACCGGCACCGCATCCGCCGC 4700
 T T L T S Q L T D S D R D R I R R
 GGCGGCTTCCCTGCCGATCTGGACGACGAGGGCATGC
 G G F L P I S D D E G M

20 The *NheI-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 13 of rapamycin is shown below.

GCATGCGCTGTACGAGGGCACGGCGACCGGAAGTCCGTGGTGGT 50
 M R L Y E A A R R T G S P V V V
 25 GCGGCCGCGCTCGACGACGCGCCGGACGTGCCGCTGCTGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCTGCCGCCGTCCGGGAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGAGCGCGGCCACGCCCTCCCTCGCGTTCG 200
 R S P C C P T T S A P T P P S R S
 30 TCCCTGGAACAGCACGCCACCGTCTGGCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACCTGGCATCGACTCGCTACCGCGG 300
 P A T T T F K E L G I D S L T A
 TCCAGCTGCGCACCGCTGACCGACGGCGACCGCGTACGCCCAACGCC 350
 35 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCGCGCGCTGCCCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCGCGCCCGTGCACGCCGGACCGCGGCA 450
 D E L A G T R A P V A A R T A A
 40 CCGCGGCCGCGCACGAACCGCTGGCGATCGTGGCATGGCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCGGGCGGGGCTCGCTGCCACAGGAGCTGTGGCGTCTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 45 CGGCACCGACGCCATACGGAGTTCCCGCGGACCGCGGCTGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCGACGGCATGGCAAGACCTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCCTCGACGGTGCACCGGCTTCGACGCCGTTCTCGG 700
 50 H G G F L D G A T G F D A A A F F G
 GATCAGCCCCCGCGAGGCCCTGCCATGGACCCGACGACACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCCTGGAGGCCTCGAAAGCCGGCATCACCCGGACGCG 800
 L E T S W E A F E S A G I T P D A
 GCGCGGGGAGCGACACCGCGCTGGTACCGGCTCGCTCTACGGGTA 850
 55 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTCGGCCGACAGGGTCGACGACCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGCCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 60 GTCACGGTCGACACCGCCCTGCTCGTCACTGGTCGCCCTGCACCAAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTGCGCTGGGCGAATGCTCGTCCGCTGGTGGCGGTG 1050

G Q S L R S G E C S L A L V G G
 TCACGGTGTGATGGCGTCGCCGGCGGATTCTCGAGTTCTCCCGGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGCGGGCGAAGGCCTCGGCCGGCGCGGACGG 1150
 5 G L A F D G R A K A F G A G A D G
 TAGCAGGCTTCGCGAGGGCGCCGGTGCCCTGGTGGTCGAGCGGCTCTCG 1200
 T S F A E G A G A L V V E R L S
 ACGCGGAGCGCCACGCCACACCGTCCCTGCCCTCGTACCGGCTCGCG 1250
 D A E R H G H T V L A L V R G S A
 10 GCTAACTCCGACGGCGCGTCAACGGCTGTGCGGCCGAACGGCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCTCATCCACCAGGCCCTCGCGAACGCCAACCGGCTCGCG 1350
 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACGCCGGTCAACGGCACGGCACCGGCTCGCGAC 1400
 15 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCGAGGCCGCTCGCGACGGTACGGACAGGACGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 GCCCCCTGCTGCTCGCTCGTGAAGTCGAACATCGGCCACGCCAGGCC 1500
 P L L L G S L K S N I G H A Q A
 20 CGTCAGGGGTGCGCCGGATCATCAAGATGGTGCAGGCCATCGGCACGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGCCAGGCCGCTGCCGCACGTCGACTG 1600
 E L P P T L H A D E P S P H V D W
 GACGGCCGGTGCCTGAGCTCTGACGTCGGCCGGCGTGGCCGGGA 1650
 25 T A G A V E L L T S A R P W P G
 CCGGTGCCCCGCCGCGCCGCTGCCGTCTCGTCTGGCGTGGCGCACG 1700
 T G R P R R A A V S S F G V S G T
 AACGCCACATCATCCTTGAGGCCAGGACGGCTCAAACGGGACCGGTCGA 1750
 N A H I I L E A G P V K T G P V E
 30 GGCAGGAGCGATCGAGGCCAGGCCGCTGAAGTAGGACCGGTCGAGGCTG 1800
 A G A I E A G P V E V G P V E A
 GACCGCTCCCCGCCGCGCCGCTAGCACCGGGCGAAGACCTCCGCTG 1850
 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCGTCTCCCGAGGCACTCGACGAGCAGATCGGGCGCCT 1900
 35 L V S A R S P E A L D E Q I G R L
 GCGCGCCTATCTGACACCAGGCCGGCGTCGACCGGGCGCGTGGCGC 1950
 R A Y L D T G P G V D R A A V A
 AGACACTGCCCGGCTACGCACTTCACCCACCGGGCGTACTGCTCGGG 2000
 Q T L A R R T H F T H R A V L L G
 40 GACACCGTATCGGCCCTCCCCCGGGACAGGCCGACGAACCTCGTCTT 2050
 D T V I G A P P A D Q A D E L V F
 CGTCTACTCCGGTCAGGGCACCCAGCATCCCGCATGGCGAGCAGCTAG 2100
 V Y S G Q G T Q H P A M G E Q L
 CCGATTCTGCGTGGTGTCTGCCGAGCGGATGGCGAGTGTGGCGCGCG 2150
 45 A D S S V V F A E R M A E C A A A
 TTGCGCGAGTTCTGGACTGGATCTGTTACGGTTCTGGATGATCCGGC 2200
 L R E F V D W D L F T V L D D P A
 GGTGGTGGACCGGGTTGATGTGGTCCAGCCCCTTCTGGCGATGATGG 2250
 V V D R V D V V Q P A S W A M M
 50 TTTCCCTGGCCCGGGTGTGGCAGGCCGGCGATGGCGAGCAGCTG 2300
 V S L A A V W Q A A G V R P D A V
 ATCGGCCATTGCAAGGGTGAGATCGCCGCAGCTGTGTGGCGGGTGC 2350
 I G H S Q G E I A A A A C V A G A V
 GTCACTACCGCAGGCCGCCGGATCGTGACCTTGCAGGCCAGGGCATCG 2400
 55 S L R D A A R I V T L R S Q A I
 CCCGGGGCTGGCGGGCGGGCGCATGGCATCCGTGCCCTGCCCGCG 2450
 A R G L A G R G A M A S V A L P A
 CAGGATGTCGAGCTGGTCGACGGGGCTGGATCGCCGCCAACACGGGCC 2500
 Q D V E L V D G A W I A A H N G P
 60 CGCCTCCACCGTGTGCGGGCACCCCGGAAGCGGTGACCATGTCCTCA 2550
 A S T V I A G T P E A V D H V L
 CCGCTCATGAGGCACAAGGGGTGCGGGTGCAGGCCGATCACCGTGC 2600
 T A H E A Q G V R V R R I T V D Y
 GCCTCGCACACCCCGCACGTCGAGCTGATCCGCGACGAACACTCGACAT 2650

A S H T P H V E L I R D E L L D I
 CACTAGCGACAGCAGCTCGCAGACCCCGCTCGTGCCTGGCTGTCGACCG 2700
 T S D S S S Q T P L V P W L S T
 5 TGGACGGCACCTGGTCGACAGCCCGCTGGACGGGGAGTACTGGTACCGG 2750
 V D G T W V D S P L D G E Y W Y R
 AACCTCGTCAACCGGTGGTTCCACCCCGCCGTCAAGCCAGTTGCAGGC 2800
 N L R E P V G F H P A V S Q L Q A
 CCAGGGCGACACCGTGTTCGAGGTCAAGCGCCAGCCCGGTGTGTC 2850
 Q G D T V F V E V S A S P V L L
 10 AGGCATGGACGACGATGTCGTACGGTTGCCACGCTGCGTCGTGACGAC 2900
 Q A M D D D V V T V A T L R R D D
 GGCAGGCCACCCGGATGCTCACCGCCCTGGCACAGGCCATGTCACGG 2950
 G D A T R M L T A L A Q A Y V H G
 CGTCACCGTCGACTGGCCCGCCATCCTCGGCACCACAAACCCGGTAC 3000
 V T V D W P A I L G T T T T R V
 TGGACCTCCGACCTACGCCCTCCAACACCAAGCGGTACTGGCTCGAGTCG 3050
 L D L P T Y A F Q H Q R Y W L E S
 GCTCCCCCGGCACGGCCGACTCGGGCACCCCGTCTCGGCACCGGAGT 3100
 A P P A T A D S G H P V L G T G V
 20 CGCCGTCGCCGGTGCACGGGGCGGGTGTTCACGGTCCCGTGCACGG 3150
 A V A G S P G R V F T G P V P A
 GTGCGGACCGCGCGGGTGTTCATCGCGAACCTGGCGTCGCCGCCGAC 3200
 G A D R A V F I A E L A L A A A D
 GCCACCGACTCGGCCACGGTCAACAGCTCGACGTCACTCCGTGCCGG 3250
 25 A T D C A T V E Q L D V T S V P G
 CGGATCCGCCCGGGCAGGGCACCGCGCACGACCTGGTCGATGAACCCG 3300
 G S A R G R A T A Q T W V D E P
 CGCCGACGGCGGCCGCTTCACCGTCCACACCCCGCTGGCGACGCC 3350
 A A D G R R R F T V H T R V G D A
 30 CCGTGGACGCTGCACGCCGAGGGGTTCTCGCCCGCGCGTGCCCCA 3400
 P W T L H A E G V L R P G R V P Q
 GCCCGAAGCCCGTCGACACCGCTGGCCCCCGCCGGCGCGTGCCCCG 3450
 P E A V D T A W P P P G A V P A
 ACGGGCTGCCGGGGCGTGGCGACGGGACCGAGGTCTCGTCAAGCC 3500
 35 D G L P G A W R R A D Q V F V E A
 GAAGTCGACAGCCCTGACGGCTTGTGGCACACCCGACCTGCTCGACGC 3550
 E V D S P D G F V A H P D L L D A
 GGTCTTCTCCGCGGTGGCGACGGGAGCCGCCAGCCGACCGGATGGCG 3600
 V F S A V G D G S R Q P T G W R
 40 ACCTCGCGGTGCACCGCTCGGACGCCACCGTGCTGCGCGCTCACC 3650
 D L A V H A S D A T V L R A C L T
 CGCCGCGACAGTGGTGTGGAGCTCGCCGCCCTCGACGGTGCCGGAAT 3700
 R R D S G V V E L A A F D G A G M
 GCCGGTGCTCACCGCGGAGTCGGTGACGCTGGCGAGGTGCGCGCAG 3750
 45 P V L T A E S V T L G E V A S A
 GCGGATCCGACGAGTCGGACGGTCTGCTTGGCTTGAGTGGTTGCCGGT 3800
 G G S D E S D G L L R L E W L P V
 GCGGAGGCCCACTACGACGGTGGCGACGAGCTGCCGAGGGCTACACCC 3850
 A E A H Y D G A D E L P E G Y T L
 50 CATCACCGCCACACACCCGACGACCCGACGACCCACCAACCCCA 3900
 I T A T H P D D P D P T N P H
 ACACACCCACACGCACCCACACACAAACACACCGCTCTCACGCCCTC 3950
 N T P T R T H T Q T T R V L T A L
 CAACACCACCTCATCACCAACCACACCCCATCGTCCACACCACAC 4000
 55 Q H H L I T T N H T L I V H T T T
 CGACCCCCCAG3CGCCGCCGTACCGGCCCTACCCGCACCGCACAAACG 4050
 D P P G A A V T G L T R T A Q N
 AACACCCCCGGCCGCATCCACCTCATCGAAACCCACCAACCCCA 4100
 E H P G R I H L I E T H H P H T P
 60 CTCCCCCTCACCCAACTCACCAACCCCTCCACCAACCCACCTACGCC 4150
 L P L T Q L T T L H Q P H L R L T
 CAACAAACACCCCTCCACACCCCCCACCTCACCCCATCACCAACCA 4200
 N N T L H T P H L T P I T T H H
 ACACCAACCAACCACCCCCAACACCCACCCCTCAACCCCAACCGCC 4250

```

N T T T T P N T P P L N P N H A
ATCCTCATACCGGGCTCCGGCACCTCGCCGGATCCTCGCCCCGCCA 4300
I L I T G G S G T L A G I L A R H
CCTCAACCACCCCCACACCTACCTCCTCTCCCGCACACCACCCCCCA 4350
5 L N H P H T Y L L S R T P P P P
CCACACCCGGCACCCACATCCCCCTGCGACCTCACCGAACCCACCCAAATC 4400
T T P G T H I P C D L T D P T Q I
ACCCAAGCCCTCACCCACATACCAACCCCTCACCGGCATCTCCACAC 4450
T Q A L T H I P Q P L T G I F H T
10 CGCCGCCACCCCTCGACGACGCCACCCCTACCAACCTACCCCCAACACC 4500
A A T L D D A T L T N L T P Q H
TCACCACCAACCCCTCCAACCCAAAGCCGACGCCGCTGGCACCTCCACAC 4550
L T T L Q P K A D A A W H L H H
CACACCCAAAACCAACCCCTCACCCACTTCGTCTACTCCAGCGCCGC 4600
15 H T Q N Q P L T H F V L Y S S A A
CGCCACCCCTCGGCAGCCCCGGCCAAGCCAACCTACGCCGCCAACGCCT 4650
A T L G S P G Q A N Y A A A N A
TCCTCGACGCCCTGCCACCCACCGCCACACCCAAGGACAACCCGCCACC 4700
20 F L D A L A T H R H T Q G Q P A T
ACCATGCCCTGGGCATGTGGCACACCACACTCACCAACTGCCACT 4750
T I A W G M W H T T T L T S Q L
CACCGACAGCGACCGCGACCGCATCCGCCGGCGCTCTGCCGATCT 4800
T D S D R D R I R R G G F L P I
25 CGGACGACGGAGGGCATGC
S D D E G M

```

Example 3

Recombinant PKS Genes for 13-desmethoxy FK-506 and FK-520

The present invention provides a variety of recombinant PKS genes in addition to those described in Examples 1 and 2 for producing 13-desmethoxy FK-506 and FK-520 compounds. This Example provides the construction protocols for recombinant FK-520 and FK-506 (from *Streptomyces* sp. MA6858 (ATCC 55098), described in U.S. Patent Nos. 5,116,756, incorporated herein by reference) PKS genes in which the module 8 AT coding sequences have been replaced by either the *rapAT3* (the AT domain from module 3 of the rapamycin PKS), *rapAT12*, *eryAT1* (the AT domain from module 1 of the erythromycin (DEBS) PKS), or *eryAT2* coding sequences. Each of these constructs provides a PKS that produces the 13-desmethoxy-13-methyl derivative, except for the *rapAT12* replacement, which provides the 13-desmethoxy derivative, i.e., it has a hydrogen where the other derivatives have methyl.

Figure 7 shows the process used to generate the AT replacement constructs. First, a fragment of ~4.5 kb containing module 8 coding sequences from the FK-520 cluster of ATCC 14891 was cloned using the convenient restriction sites *SacI* and *SphI* (Step A in Figure 7). The choice of restriction sites used to clone a 4.0 - 4.5 kb fragment comprising module 8 coding sequences from other FK-520 or FK-506 clusters can be different depending on the DNA sequence, but the overall scheme is identical. The unique *SacI* and *SphI* restriction sites at the ends of the FK-520 module 8 fragment were then changed to unique *Bgl* II and *Nsi* I sites by ligation to synthetic linkers (described in

the preceding Examples, see Step B of Figure 7). Fragments containing sequences 5' and 3' of the AT8 sequences were then amplified using primers, described above, that introduced either an *AvrII* site or an *NheI* site at two different KS/AT boundaries and an *XhoI* site at the AT/DH boundary (Step C of Figure 7). Heterologous AT domains from 5 the rapamycin and erythromycin gene clusters were amplified using primers, as described above, that introduced the same sites as just described (Step D of Figure 7). The fragments were ligated to give hybrid modules with in-frame fusions at the KS/AT and AT/DH boundaries (Step E of Figure 7). Finally, these hybrid modules were ligated 10 into the *BamHI* and *PstI* sites of the KC515 vector. The resulting recombinant phage were used to transform the FK-506 and FK-520 producer strains to yield the desired recombinant cells, as described in the preceding Examples.

The following table shows the location and sequences surrounding the engineered site of each of the heterologous AT domains employed. The FK-506 hybrid construct was used as a control for the FK-520 recombinant cells produced, and a similar FK-520 15 hybrid construct was used as a control for the FK-506 recombinant cells.

Heterologous AT	Enzyme	Location of Engineered Site
FK-506 AT8 (hydroxymalonyl)	<i>AvrII</i>	GGCCGT <u>ccgcgc</u> CGTGC GGCGGT TCGTCGTT G R P R R A A V S S F
	<i>NheI</i>	ACCCAGCATCCCGC GATGGGT GAGCG <u>gctcgcc</u> T Q H P A M G E R L A
	<i>XhoI</i>	TACGCC TTCC CAGCGCGGCC TACTGG <u>atcgaa</u> Y A F Q R R P Y W I E
rapamycin AT3 (methylmalonyl)	<i>AvrII</i>	GACCGG <u>ccccat</u> CGGGCGGGCGTGT CGTC CCTTC D R P R R A G V S S F
	<i>NheI</i>	TGGCAGTGGCTGGGATGGGCAGTGC <u>cctgcgg</u> W Q W L G M G S A L R
	<i>XhoI</i>	TACGCC TTCC AAACACCAGCGGT TACTGG <u>gtcgaa</u> Y A F Q H Q R Y W V E
rapamycin AT12 (malonyl)	<i>AvrII</i>	GGCCGA <u>gcgcgc</u> CGGGCAGGCC GT TCGTCCTTC G R A R R A G V S S F
	<i>NheI</i>	TCGCAGCGTGTGGCATGGGTGAGGA <u>actggcc</u> S Q R A G M G E E L A
	<i>XhoI</i>	TACGCC TTCC CAGCACCA GCGCT <u>ACTGG</u> <u>gtcgaa</u> Y A F Q H Q R Y W L E
DEBS AT1 (methylmalonyl)	<i>AvrII</i>	GCGCGA <u>ccgcgc</u> CGGGCGGGGTCTCGTCGTT A R P R R A G V S S F
	<i>NheI</i>	TGGCAGTGGCGGGCATGGCC GT CG <u>Acctgctc</u> W Q W A G M A V D L L
	<i>XhoI</i>	TACCCGTTCCAGCGCGAGCGCGT CTGG <u>gtcgaa</u> Y P F Q R E R V W L E
DEBS AT2 (methylmalonyl)	<i>AvrII</i>	GACGGG <u>gtgcgc</u> CGGGCAGGTGT CG GGCGTTC D G V R R A G V S A F
		GCCCAGTGGGAAGGCATGGCGGGG <u>Aattat</u> GG

	<i>NheI</i>	A Q W E G M A R E L L TATCCTTCAGGGCAAGCGGTTCTGG <u>ctgcctg</u>
	<i>XbaI</i>	Y P F Q G K R F W L L

The sequences shown below provide the location of the KS/AT boundaries chosen in the FK-520 module 8 coding sequences. Regions where *AvrII* and *NheI* sites were engineered are indicated by lower case and underlining.

5 CCGGCGCCGTCGAACGTGCTGACGTCGGCCGGCGTGGCCCGAGACCGACCGccacggC
 A G A V E L L T S A R P W P E T D R P R
 GTGCCGCCGTCCTCGTTGGGTGAGCGGCACCAACGCCACGTACCTGGAGGCCG
 R A A V S S F G V S G T N A H V I L E A
 GACCGGTAACGGAGACGCCGCGCATGCCCTCCGGTGCACCTCCCTGCTGGTGTGG
 G P V T E T P A A S P S G D L P L L V S
 10 CACGCTCACCGGAAGCGCTCGACGAGCAGATCCGCCACTGCGCGCTACCTGGACACCA
 A R S P E A L D E Q I R R L R A Y L D T
 CCCCCGACGTCGACCGGGTGGCGTGGCACAGACGCTGGCCGGCACACACTCGCCC
 T P D V D R V A V A Q T L A R R T H F A
 15 ACCGCGCCGTCGCTCGGTGACACCGTCATCACCAACACCCCGCGGACCGGCCGACG
 H R A V L L G D T V I T T P P A D R P D
 AACTCGTCTCGTCATCCGGCAGGGCACCCAGCATCCCGCATGGCGAGCAgctcg
 E L V F V Y S G Q G T Q H P A M G E Q L
 CCGCCGCCATCCCGTGTCCGCGACGCCATGAAGCGCTCCGCCGCTTGACAAACC
 A A A H P V F A D A W H E A L R R L D N
 20

The sequences shown below provide the location of the AT/DH boundary chosen in the FK-520 module 8 coding sequences. The region where an *XhoI* site was engineered is indicated by lower case and underlining.

25 TCCTCGGGCTGGTCACGGCACGACGCGATGTGCCCGTACCGCTTCCAACGGCGGC
 I L G A G S R H D A D V P A Y A F Q R R
 ACTACTGGatcgagTCCGGCACGCCGGCGATCCGACGCCGGACCCGTGCTGGGCT
 H Y W I E S A R P A A S D A G H P V L G

The sequences shown below provide the location of the KS/AT boundaries chosen in the FK-506 module 8 coding sequences. Regions where *AvrII* and *NheI* sites were engineered are indicated by lower case and underlining.

30 TCGGCCAGGCCGTGGCGCGGACCGGGCGTccgcgcCGTGCGGCGTCTCGTCTCGGG
 S A R P W P R T G R P R R A A V S S F G
 35 GTGAGCGGCCACCAACGCCACATCATCTGGAGGCCGACCCGACCAGGAGGCCGTCG
 V S G T N A H I I L E A G P D Q E E P S
 GCAGAACGGCCGGTACCTCCCGTCTCGTGTGGCACGGTCCCGGAGGCACTGGAC
 A E P A G D L P L L V S A R S P E A L D
 GAGCAGATCGGGCGCTCGCGACTATCTCGACGCCGCCCCGGCGTGGACCTGGCGGCC
 E Q I G R L R D Y L D A A P G V D L A A
 40 GTGGCGGGACACTGCCACCGTACGCACTCTCCACCGCGCCGTACTGCTCGGTGAC
 V A R T L A T R T H F S H R A V L L G D
 ACCGTCATACCGCTCCCCCGTGGAACAGCCGGCGAGCTCGTCTCGTCACTCGGGA
 T V I T A P P V E Q P G E L V F V Y S G
 45 CAGGGCACCCAGCATCCCGCATGGGTGAGCGgctcgCGCAGCCTCCCGTGTTCGCC
 Q G T Q H P A M G E R L A A A F P V F A
 GACCCGGACGTACCCGCCCTACGCCCTCCAGCGGGGCCACTGGATCGAGTCGCCG
 D P D V P A Y A F Q R R P Y W I E S A P

The sequences shown below provide the location of the AT/DH boundary chosen in the FK-506 module 8 coding sequences. The region where an *XhoI* site was engineered is indicated by lower case and underlining.

GACCCGGACGTACCCGCCCTACGCCCTCCAGCGGGGCCACTGGatcgagTCCGCC
 D P D V P A Y A F Q R R P Y W I E S A P

Example 4Replacement of Methoxyl with Hydrogen or Methyl at C-15 of FK-506 and FK-520

The methods and reagents of the present invention also provide novel FK-506 and FK-520 derivatives in which the methoxy group at C-15 is replaced by a hydrogen or methyl. These derivatives are produced in recombinant host cells of the invention that express recombinant PKS enzymes that produce the derivatives. These recombinant PKS enzymes are prepared in accordance with the methodology of Examples 1 and 2, with the exception that AT domain of module 7, instead of module 8, is replaced. Moreover, the present invention provides recombinant PKS enzymes in which the AT domains of both modules 7 and 8 have been changed. The table below summarizes the various compounds provided by the present invention.

	Compound	C-13	C-15	Derivative Provided
15	FK-506	hydrogen	hydrogen	13, 15-didesmethoxy-FK-506
	FK-506	hydrogen	methoxy	13-desmethoxy-FK-506
	FK-506	hydrogen	methyl	13,15-didesmethoxy-15-methyl-FK-506
	FK-506	methoxy	hydrogen	15-desmethoxy-FK-506
	FK-506	methoxy	methoxy	Original Compound -- FK-506
20	FK-506	methoxy	methyl	15-desmethoxy-15-methyl-FK-506
	FK-506	methyl	hydrogen	13,15-didesmethoxy-13-methyl-FK-506
	FK-506	methyl	methoxy	13-desmethoxy-13-methyl-FK-506
	FK-506	methyl	methyl	13,15-didesmethoxy-13,15-dimethyl-FK-506
	FK-520	hydrogen	hydrogen	13, 15-didesmethoxy FK-520
25	FK-520	hydrogen	methoxy	13-desmethoxy FK-520
	FK-520	hydrogen	methyl	13,15-didesmethoxy-15-methyl-FK-520
	FK-520	methoxy	hydrogen	15-desmethoxy-FK-520
	FK-520	methoxy	methoxy	Original Compound -- FK-520
	FK-520	methoxy	methyl	15-desmethoxy-15-methyl-FK-520
30	FK-520	methyl	hydrogen	13,15-didesmethoxy-13-methyl-FK-520
	FK-520	methyl	methoxy	13-desmethoxy-13-methyl-FK-520
	FK-520	methyl	methyl	13,15-didesmethoxy-13,15-dimethyl-FK-520

Example 5Replacement of Methoxyl with Ethyl at C-13 and/or C-15 of FK-506 and FK-520

The present invention also provides novel FK-506 and FK-520 derivative compounds in which the methoxy groups at either or both the C-13 and C-15 positions are instead ethyl groups. These compounds are produced by novel PKS enzymes of the invention in which the AT domains of modules 8 and/or 7 are converted to ethylmalonyl specific AT domains by modification of the PKS gene that encodes the module. Ethylmalonyl specific AT domain coding sequences can be obtained from, for example, the FK-520 PKS genes, the niddamycin PKS genes, and the tylisin PKS genes. The novel PKS genes of the invention include not only those in which either or both of the AT domains of modules 7 and 8 have been converted to ethylmalonyl specific AT domains but also those in which one of the modules is converted to an ethylmalonyl specific AT domain and the other is converted to a malonyl specific or a methylmalonyl specific AT domain.

Example 6

15 Neurotrophic Compounds

The compounds described in Examples 1 - 4, inclusive have immunosuppressant activity and can be employed as immunosuppressants in a manner and in formulations similar to those employed for FK-506. The compounds of the invention are generally effective for the prevention of organ rejection in patients receiving organ transplants and 20 in particular can be used for immunosuppression following orthotopic liver transplantation. These compounds also have pharmacokinetic properties and metabolism that are more advantageous for certain applications relative to those of FK-506 or FK-520. These compounds are also neurotrophic; however, for use as neurotrophins, it is desirable to modify the compounds to diminish or abolish their immunosuppressant 25 activity. This can be readily accomplished by hydroxylating the compounds at the C-18 position using established chemical methodology or novel FK-520 PKS genes provided by the present invention.

Thus, in one aspect, the present invention provides a method for stimulating nerve growth that comprises administering a therapeutically effective dose of 18-hydroxy-FK-520. In another embodiment, the compound administered is a C-18,20-dihydroxy-FK-520 derivative. In another embodiment, the compound administered is a C-13-desmethoxy and/or C-15-desmethoxy 18-hydroxy-FK-520 derivative. In another embodiment, the compound administered is a C-13-desmethoxy and/or C-15-desmethoxy 18,20-dihydroxy-FK-520 derivative. In other embodiments, the compounds 35 are the corresponding analogs of FK-506. The 18-hydroxy compounds of the invention

can be prepared chemically, as described in U.S. Patent No. 5,189,042, incorporated herein by reference, or by fermentation of a recombinant host cell provided by the present invention that expresses a recombinant PKS in which the module 5 DH domain has been deleted or rendered non-functional.

5 The chemical methodology is as follows. A compound of the invention (~200 mg) is dissolved in 3 mL of dry methylene chloride and added to 45 μ L of 2,6-lutidine, and the mixture stirred at room temperature. After 10 minutes, tert-butyldimethylsilyl trifluoromethanesulfonate (64 μ L) is added by syringe. After 15 minutes, the reaction mixture is diluted with ethyl acetate, washed with saturated bicarbonate, washed with 10 brine, and the organic phase dried over magnesium sulfate. Removal of solvent *in vacuo* and flash chromatography on silica gel (ethyl acetate: hexane (1:2) plus 1% methanol) gives the protected compound, which is dissolved in 95% ethanol (2.2 mL) and to which is added 53 μ L of pyridine, followed by selenium dioxide (58 mg). The flask is fitted with a water condenser and heated to 70°C on a mantle. After 20 hours, the mixture is 15 cooled to room temperature, filtered through diatomaceous earth, and the filtrate poured into a saturated sodium bicarbonate solution. This is extracted with ethyl acetate, and the organic phase is washed with brine and dried over magnesium sulfate. The solution is concentrated and purified by flash chromatography on silica gel (ethyl acetate: hexane (1:2) plus 1% methanol) to give the protected 18-hydroxy compound. This compound is 20 dissolved in acetonitrile and treated with aqueous HF to remove the protecting groups. After dilution with ethyl acetate, the mixture is washed with saturated bicarbonate and brine, dried over magnesium sulfate, filtered, and evaporated to yield the 18-hydroxy compound. Thus, the present invention provides the C-18-hydroxyl derivatives of the compounds described in Examples 1 - 4.

25 Those of skill in the art will recognize that other suitable chemical procedures can be used to prepare the novel 18-hydroxy compounds of the invention. See, e.g., Kawai *et al.*, Jan. 1993, Structure-activity profiles of macrolactam immunosuppressant FK-506 analogues, *FEBS Letters* 316(2): 107-113, incorporated herein by reference. These methods can be used to prepare both the C18-[*S*]-OH and C18-[*R*]-OH enantiomers, with 30 the *R* enantiomer showing a somewhat lower IC₅₀, which may be preferred in some applications. See Kawai *et al.*, *supra*. Another preferred protocol is described in Umbreit and Sharpless, 1977, *JACS* 99(16): 1526-28, although it may be preferable to use 30 equivalents each of SeO₂ and t-BuOOH rather than the 0.02 and 3-4 equivalents, respectively, described in that reference.

All scientific and patent publications referenced herein are hereby incorporated by reference. The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments, that the foregoing description and example is for purposes of 5 illustration and not limitation of the following claims.

Claims

1. An isolated nucleic acid that encodes a CoA ligase, a non-ribosomal peptide synthetase, or a domain of an extender module of a polyketide synthase enzyme that synthesizes FK-520.

5

2. The isolated nucleic acid of claim 1 that encodes an extender module, said module comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.

10

3. The isolated nucleic acid of claim 1 that encodes an open reading frame, said open reading frame comprising coding sequences for two or more extender modules, each extender module comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.

15

4. The isolated nucleic acid of claim 1 that encodes a gene cluster, said gene cluster comprising two or more open reading frames, each of said open reading frames comprising coding sequences for two or more extender modules, each of said extender modules comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.

20

5. The isolated nucleic acid of claim 2, wherein at least one of said domains is a domain of a module of a non-FK-520 polyketide synthase.

25

6. The isolated nucleic acid of claim 1, wherein said nucleic acid is a recombinant vector capable of replication in or integration into the chromosome of a host cell.

30

7. The isolated nucleic acid of claim 6 that is selected from the group consisting of cosmid pKOS034-120, cosmid pKOS034-124, cosmid pKOS065-M27, and cosmid pKOS065-M21.

35

8. The isolated nucleic acid of claim 5, wherein said non-FK-520 polyketide synthase is rapamycin polyketide synthase, FK-506 polyketide synthase, or erythromycin polyketide synthase.

9. A method of preparing a polyketide, said method comprising transforming a host cell with a recombinant DNA vector of claim 6, and culturing said host cell under conditions such that said polyketide synthase is produced and catalyzes synthesis of said polyketide.

5

10. The method of claim 9, wherein said host cell is a *Streptomyces* host cell.

11. The method of claim 9, wherein said polyketide is selected from the group consisting of FK-520, 13-desmethoxy-FK-520, and 13-desmethoxy-FK-506.

10

12. A recombinant host cell that expresses a recombinant polyketide synthase selected from the group consisting of: (i) an FK-520 polyketide synthase in which at least one AT domain is replaced by an AT domain of a non-FK-520 polyketide synthase; (ii) an FK-506 polyketide synthase in which at least one AT domain is replaced by an AT domain of a non-FK-506 polyketide synthase; (iii) an FK-520 polyketide synthase in which at least one DH domain has been deleted; (iv) an FK-506 polyketide synthase in which at least one DH domain has been deleted.

15

13. The recombinant host cell of claim 12 that expresses an FK-520 polyketide synthase in which an AT domain of module 8 has been replaced by an AT domain that binds malonyl CoA, methylmalonyl CoA, or ethylmalonyl CoA.

20

14. The recombinant host cell of claim 12 that expresses an FK-506 polyketide synthase in which an AT domain of module 8 has been replaced by an AT domain that binds malonyl CoA, methylmalonyl CoA, or ethylmalonyl CoA.

25

15. The recombinant host cell of claim 13, wherein a DH domain of module 5 or module 6 has been deleted.

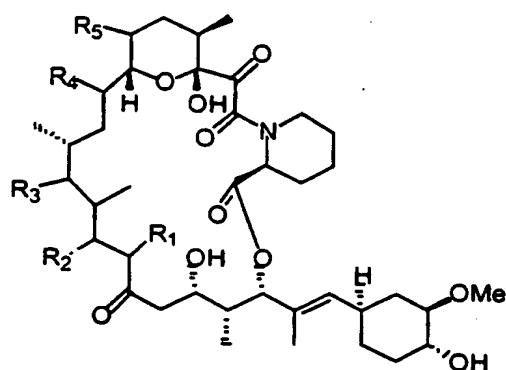
30

16. The recombinant host cell of claim 14, wherein a DH domain of module 5 or module 6 has been deleted.

17. A recombinant host cell that comprises recombinant genes coding for enzymes sufficient for synthesis of ethylmalonyl CoA or 2-hydroxymalonyl CoA.

35

18. A polyketide having the structure



5 wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen or hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506.

10

19. The polyketide of claim 18 that is 13-desmethoxy-FK-506.

20. The polyketide of claim 18 that is 13-desmethoxy-18-hydroxy-FK-520.

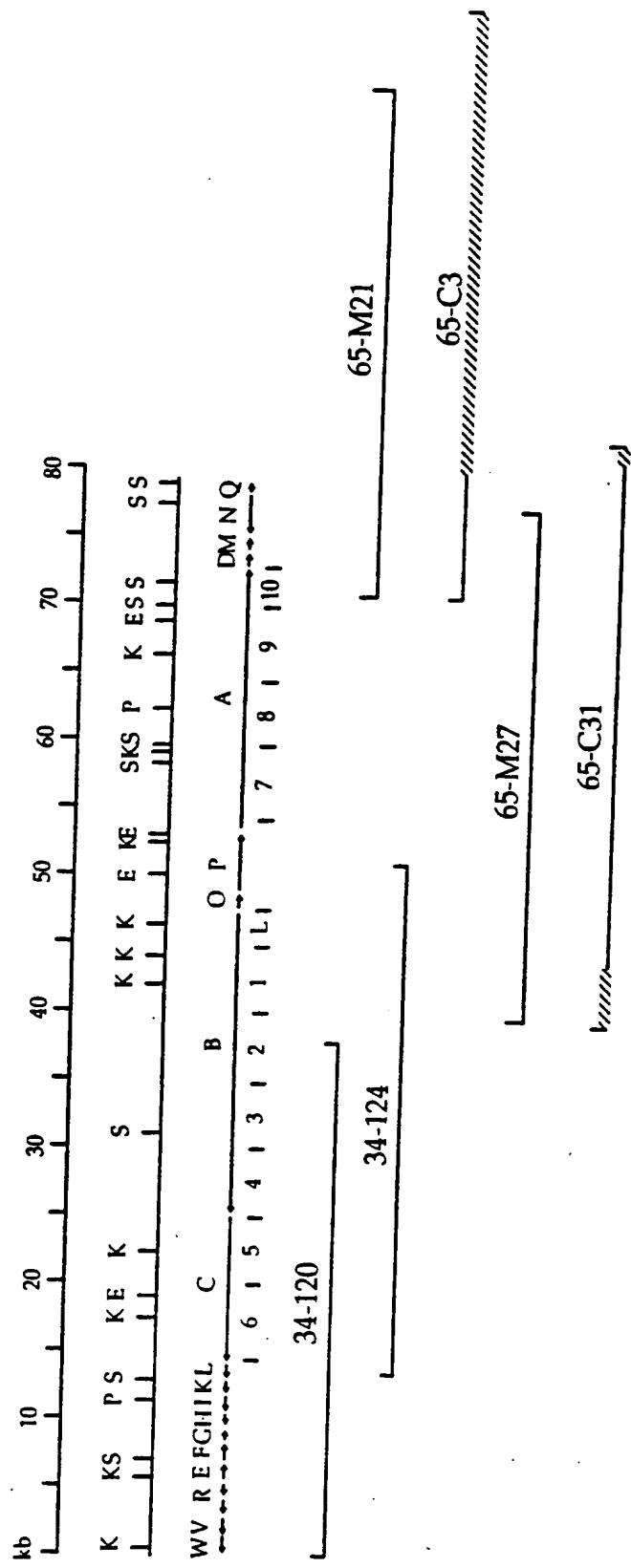


Figure 1

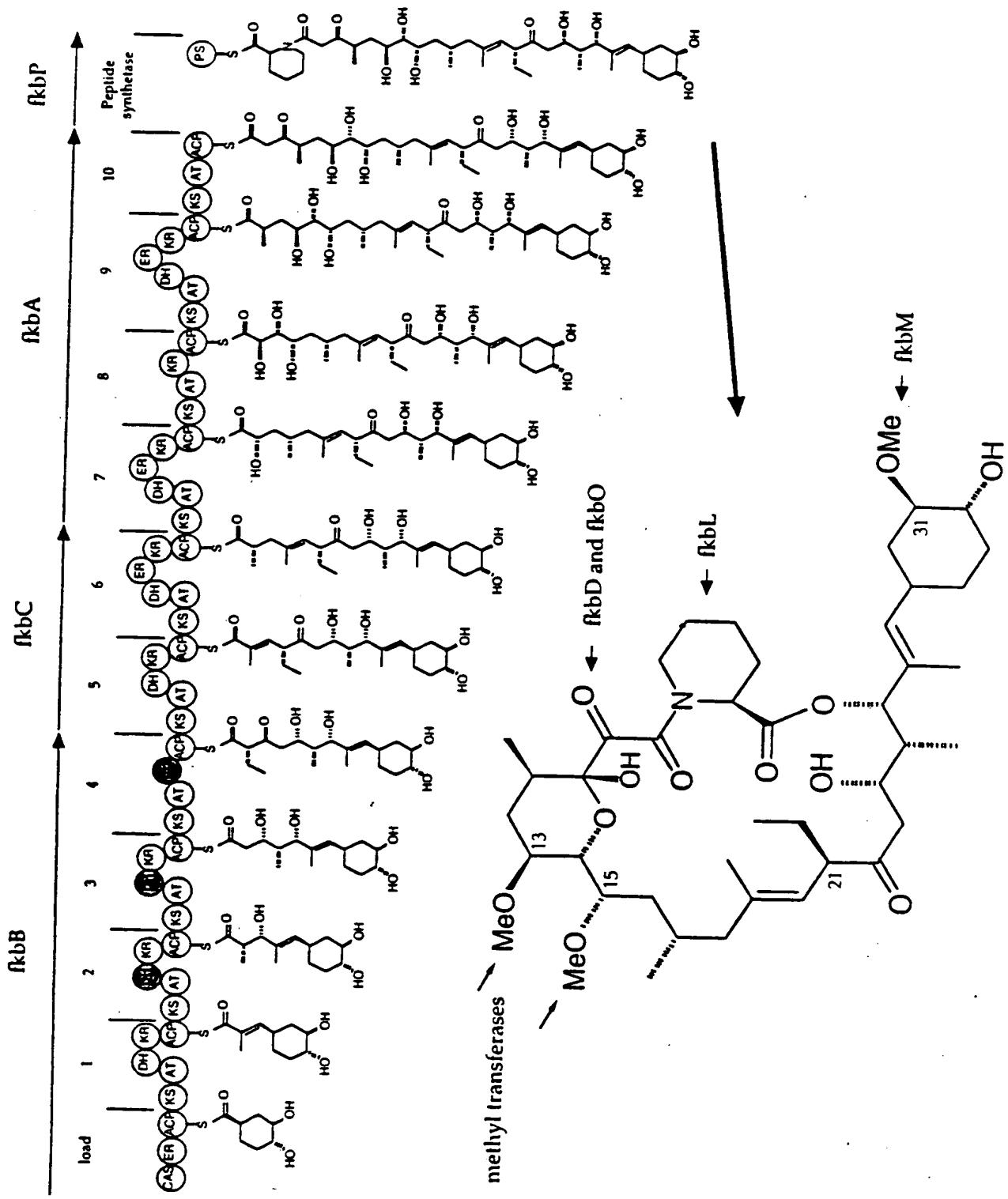


Figure 2

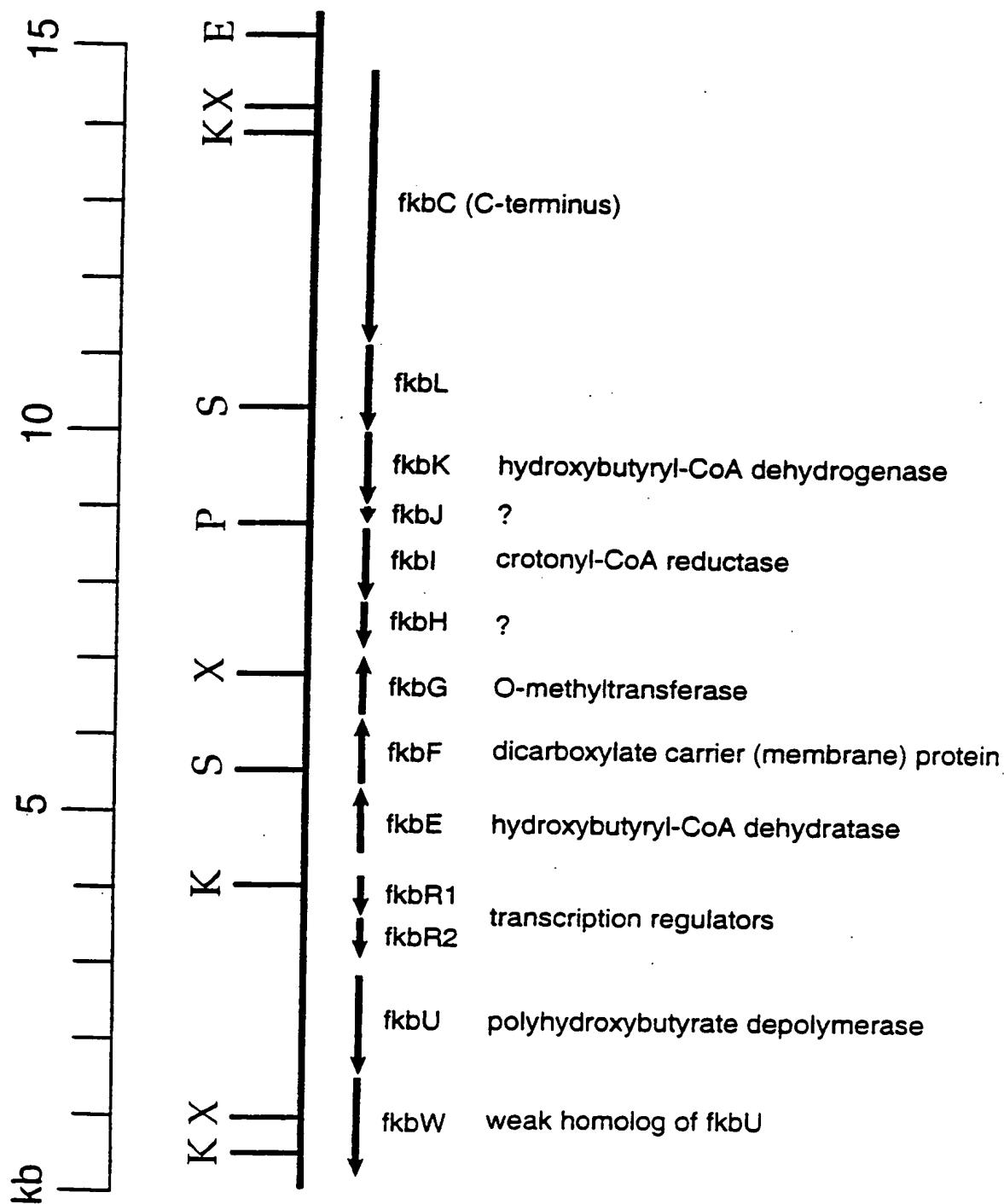


Figure 3

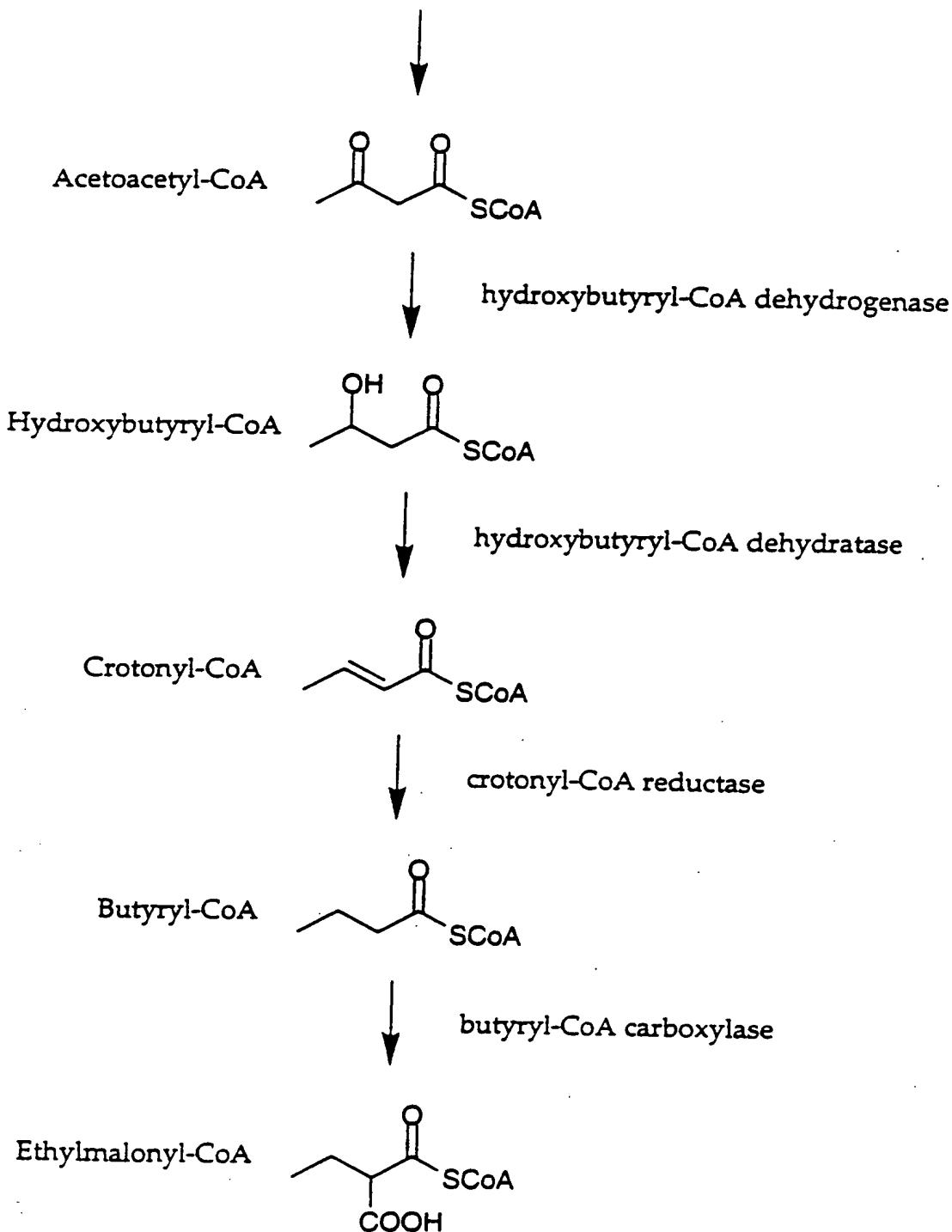


Figure 4

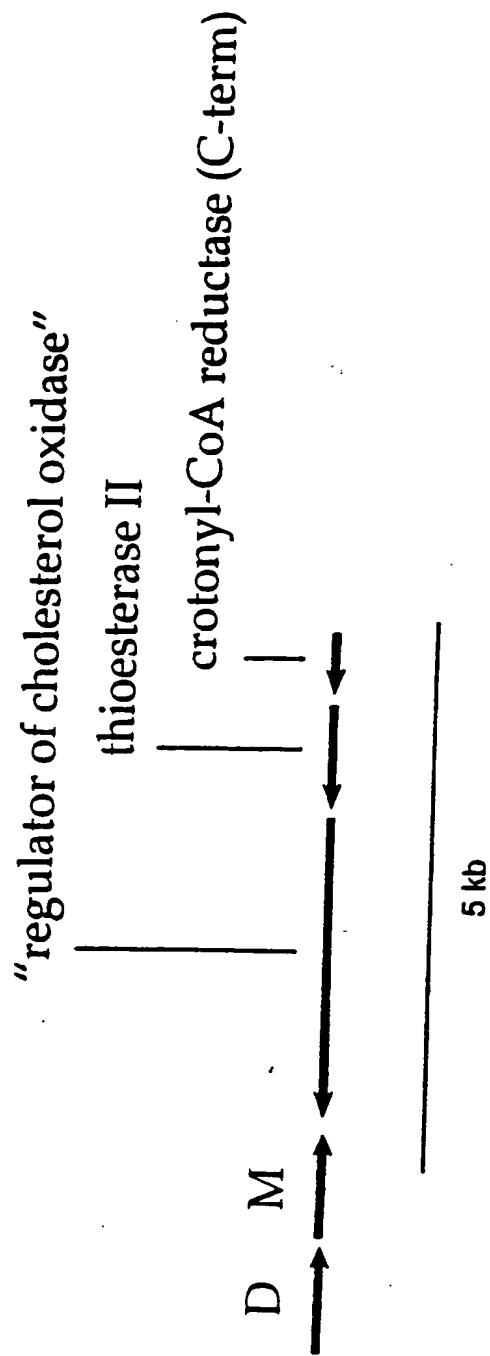


Figure 5

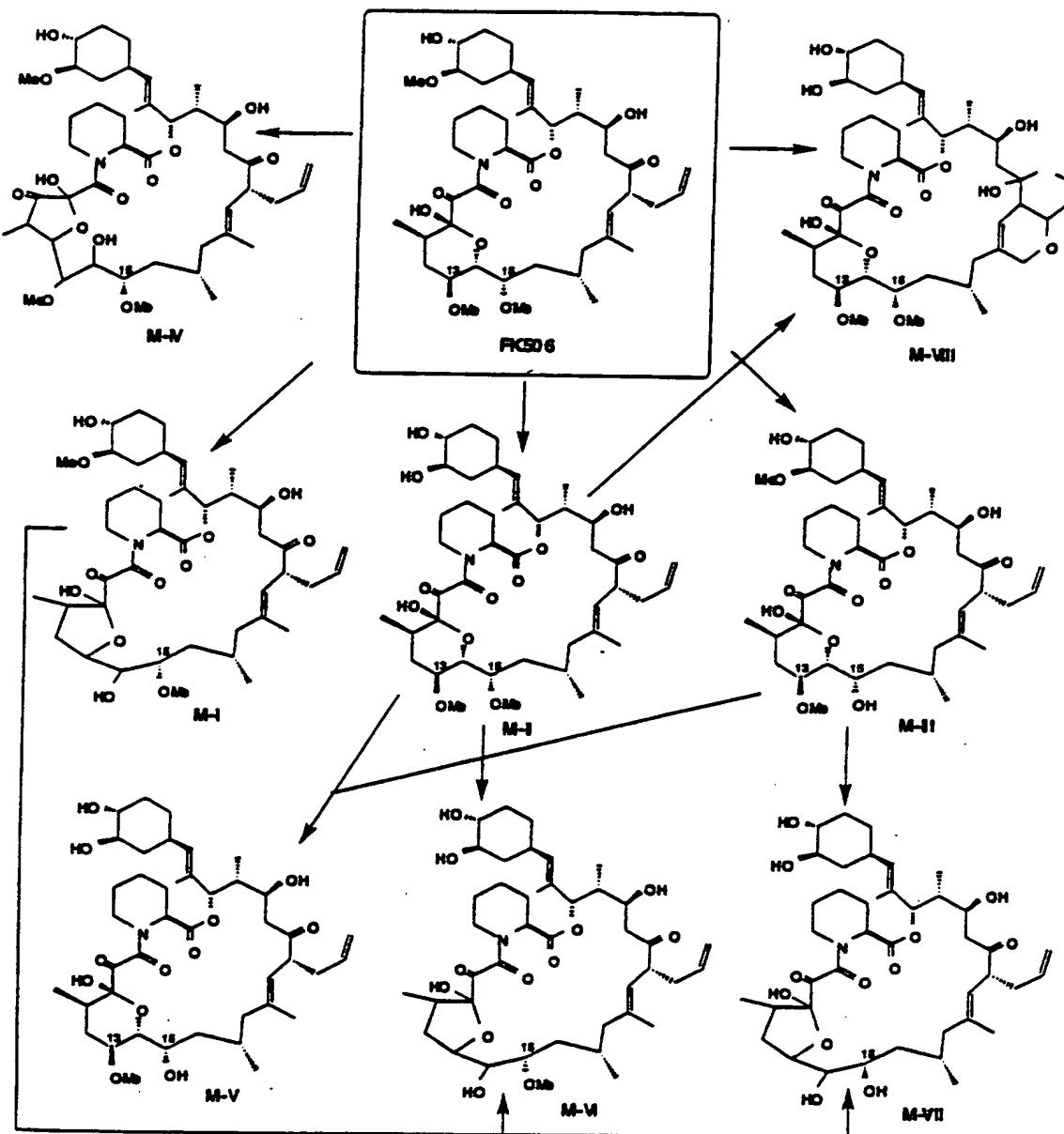


Figure 6

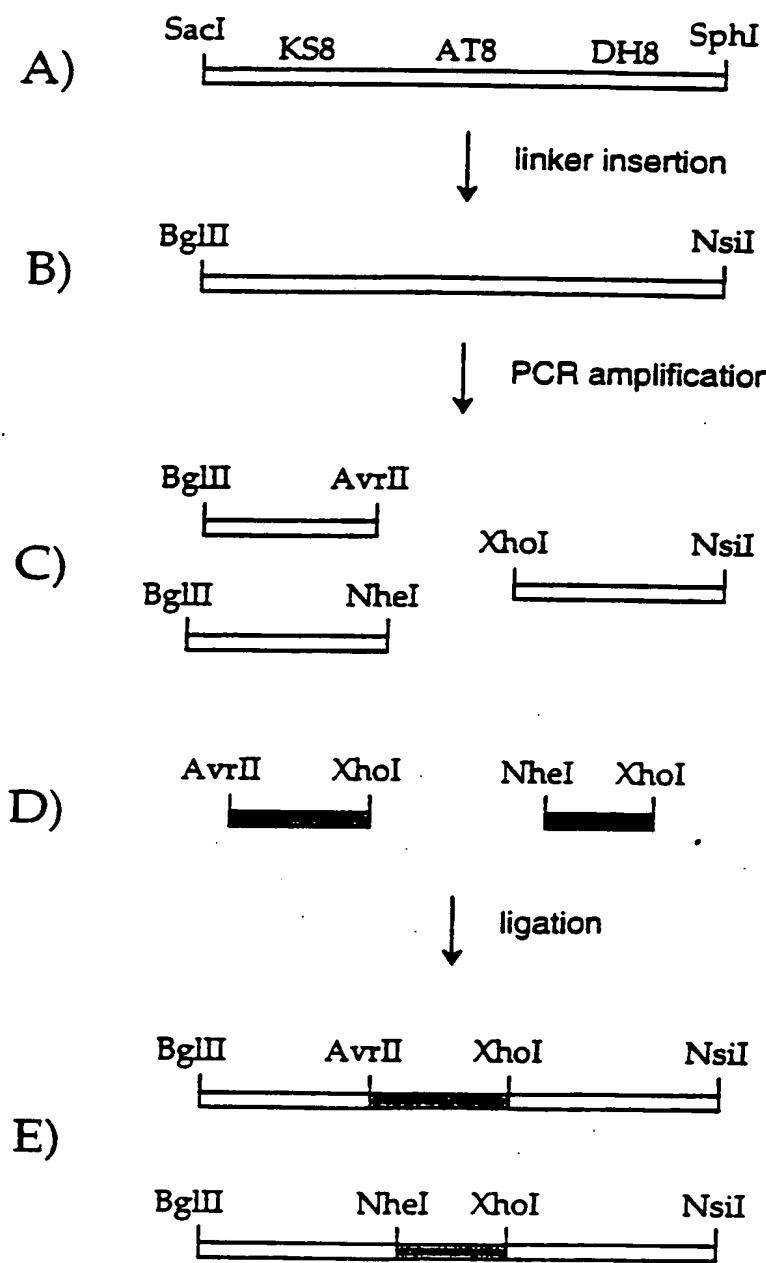


Figure 7

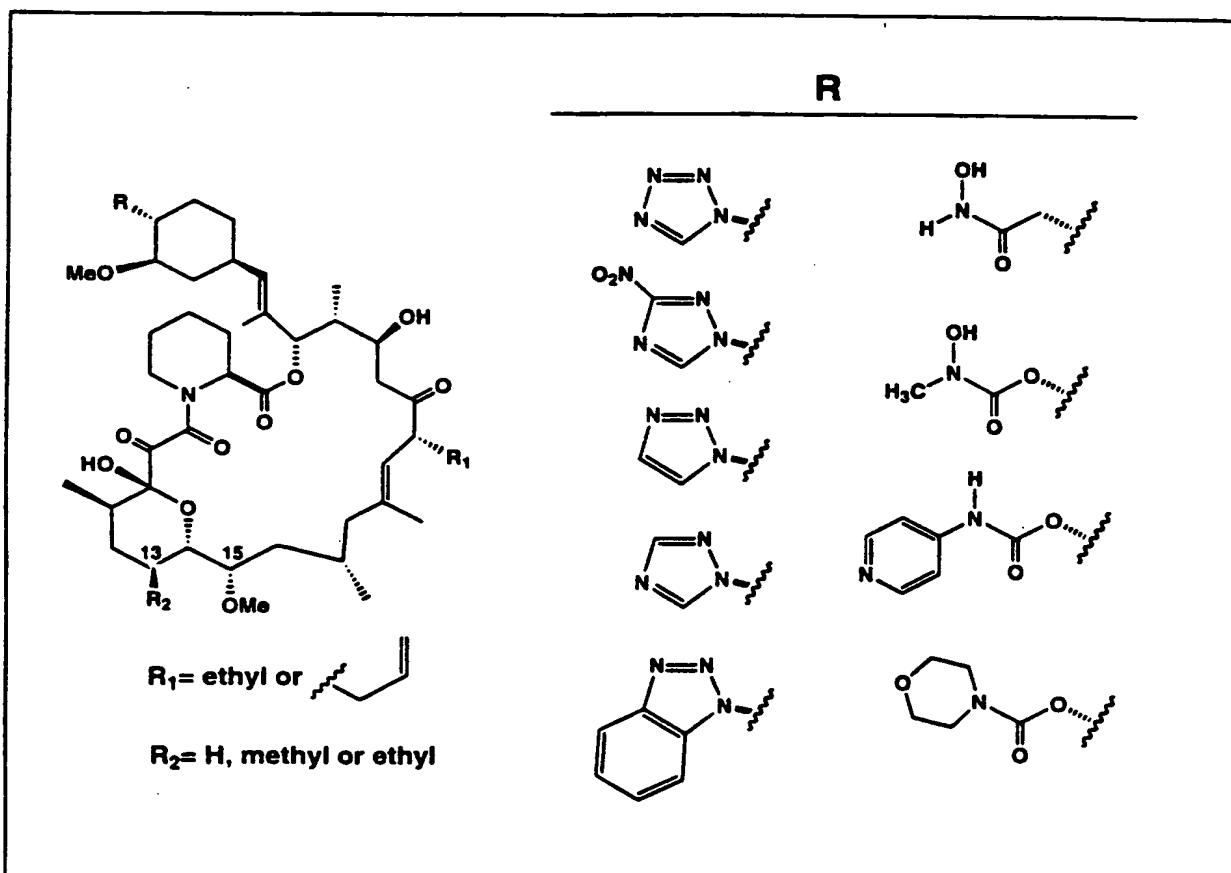


Figure 8
Part A

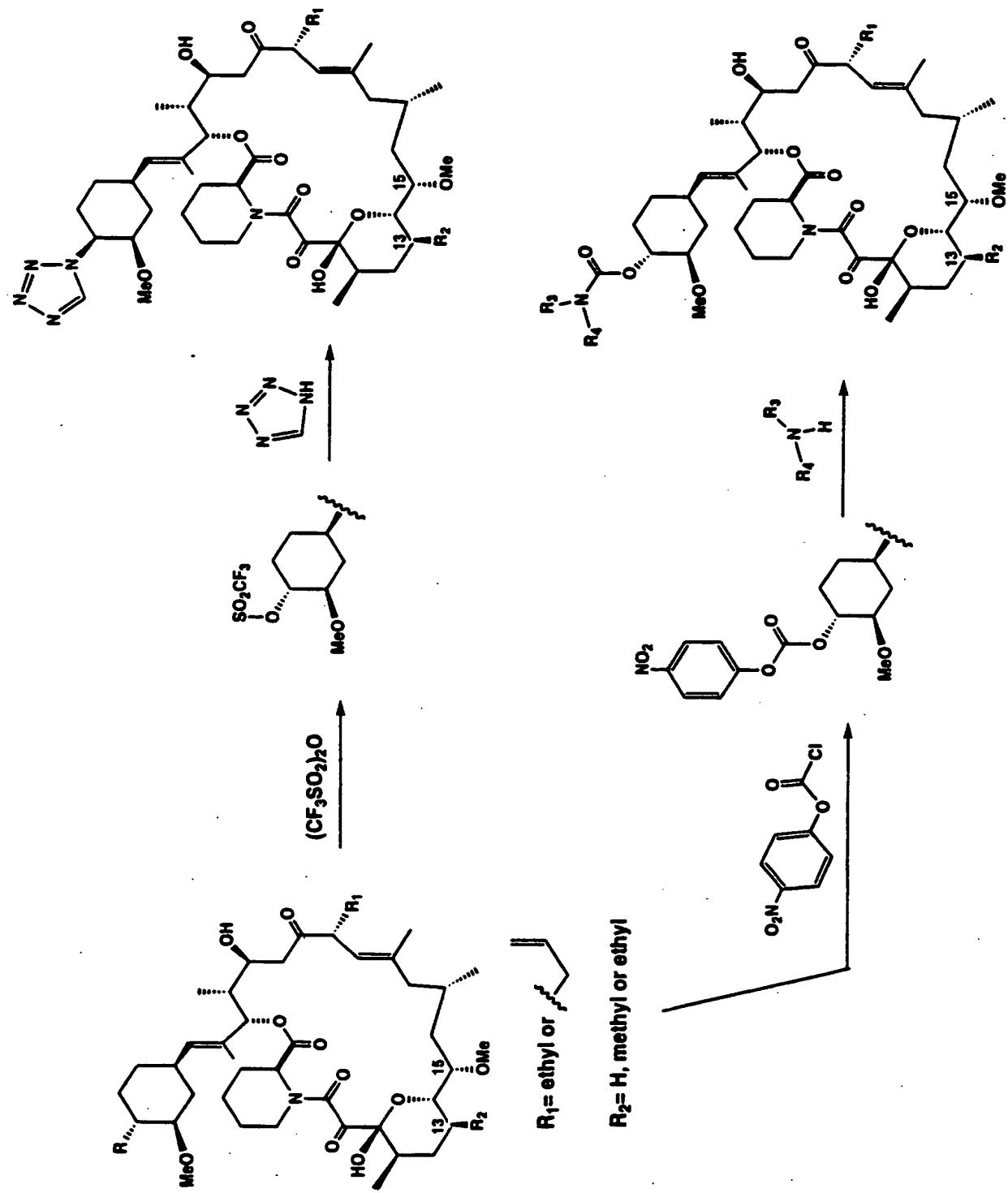


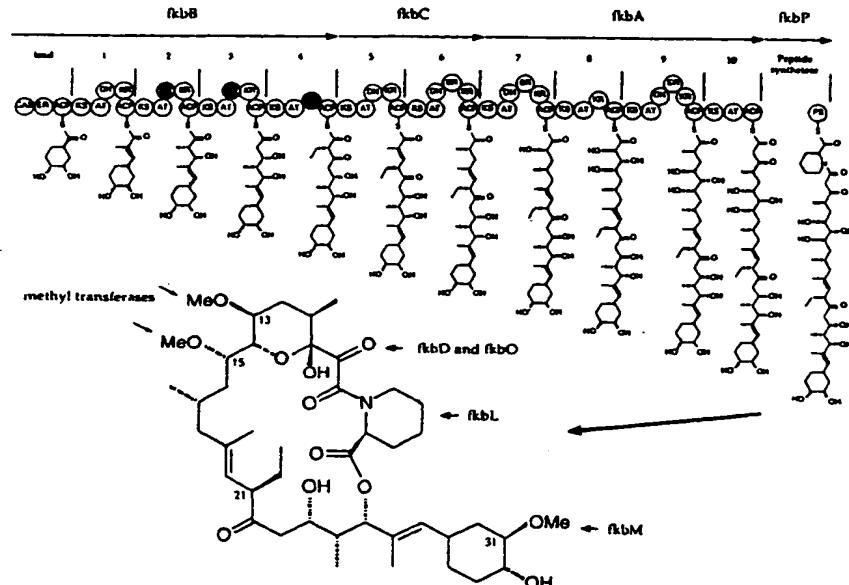
Figure 6
Part B



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/52, 15/54, 15/62, 9/10, C12P 17/18, 19/32, C07D 498/18 // (C07D 498/18, 311:00, 273:00, 211:00)		A2	(11) International Publication Number: WO 00/20601 (43) International Publication Date: 13 April 2000 (13.04.00)
(21) International Application Number: PCT/US99/22886		(74) Agents: FAVORITO, Carolyn et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).	
(22) International Filing Date: 1 October 1999 (01.10.99)			
(30) Priority Data: 60/102,748 2 October 1998 (02.10.98) US 60/123,810 11 March 1999 (11.03.99) US 60/139,650 17 June 1999 (17.06.99) US		(81) Designated States: AL, AM, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, IL, IS, JP, KG, KP, KR, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): REEVES, Christopher [US/US]; 4 East Altarinda Drive, Orinda, CA 94563 (US). CHU, Daniel [US/US]; 3767 Benton Street, Santa Clara, CA 95051 (US). KHOSLA, Chaitan [IN/US]; 740 Para Avenue, Palo Alto, CA 94306 (US). SANTI, Daniel [US/US]; 211 Belgrave Avenue, San Francisco, CA 94117 (US). WU, Kai [CN/US]; 900 Constitution Drive, Foster City, CA 94404 (US).		Published <i>Without international search report and to be republished upon receipt of that report.</i> <i>With an indication in relation to deposited biological material furnished under Rule 13bis separately from the description.</i>	

(54) Title: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS THEREFOR



(57) Abstract

Host cells comprising recombinant vectors encoding the FK-520 polyketide synthase and FK-520 modification enzymes can be used to produce the FK-520 polyketide. Recombinant DNA constructs comprising one or more FK-520 polyketide synthase domains, modules, open reading frames, and variants thereof can be used to produce recombinant polyketide synthases and a variety of different polyketides with application as pharmaceutical and veterinary products.

*(Referred to in PCT Gazette No. 35/2000, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LJ	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS THEREFOR

5

Field of the Invention

The present invention relates to polyketides and the polyketide synthase (PKS) enzymes that produce them. The invention also relates generally to genes encoding PKS enzymes and to recombinant host cells containing such genes and in which expression of such genes leads to the production of polyketides. The present invention also relates to 10 compounds useful as medicaments having immunosuppressive and/or neurotrophic activity. Thus, the invention relates to the fields of chemistry, molecular biology, and agricultural, medical, and veterinary technology.

15

Background of the Invention

Polyketides are a class of compounds synthesized from 2-carbon units through a series of condensations and subsequent modifications. Polyketides occur in many types of organisms, including fungi and mycelial bacteria, in particular, the actinomycetes. Polyketides are biologically active molecules with a wide variety of structures, and the class encompasses numerous compounds with diverse activities. Tetracycline, erythromycin, 20 epothilone, FK-506, FK-520, narbomycin, picromycin, rapamycin, spinocyn, and tylosin are examples of polyketides. Given the difficulty in producing polyketide compounds by traditional chemical methodology, and the typically low production of polyketides in wild-type cells, there has been considerable interest in finding improved or alternate means to produce polyketide compounds.

25

This interest has resulted in the cloning, analysis, and manipulation by recombinant DNA technology of genes that encode PKS enzymes. The resulting technology allows one to manipulate a known PKS gene cluster either to produce the polyketide synthesized by that PKS at higher levels than occur in nature or in hosts that otherwise do not produce the polyketide. The technology also allows one to produce molecules that are structurally related to, but distinct from, the polyketides produced from known PKS gene clusters. See, e.g., PCT publication Nos. WO 93/13663; 95/08548; 96/40968; 97/02358; 98/27203; and 30 98/49315; United States Patent Nos. 4,874,748; 5,063,155; 5,098,837; 5,149,639; 5,672,491; 5,712,146; 5,830,750; and 5,843,718; and Fu *et al.*, 1994, *Biochemistry* 33:

9321-9326; McDaniel *et al.*, 1993, *Science* 262: 1546-1550; and Rohr, 1995, *Angew. Chem. Int. Ed. Engl.* 34(8): 881-888, each of which is incorporated herein by reference.

Polyketides are synthesized in nature by PKS enzymes. These enzymes, which are complexes of multiple large proteins, are similar to the synthases that catalyze condensation of 2-carbon units in the biosynthesis of fatty acids. PKSs catalyze the biosynthesis of polyketides through repeated, decarboxylative Claisen condensations between acylthioester building blocks. The building blocks used to form complex polyketides are typically acylthioesters, such as acetyl, butyryl, propionyl, malonyl, hydroxymalonyl, methylmalonyl, and ethylmalonyl CoA. Other building blocks include amino acid like acylthioesters. PKS enzymes that incorporate such building blocks include an activity that functions as an amino acid ligase (an AMP ligase) or as a non-ribosomal peptide synthetase (NRPS). Two major types of PKS enzymes are known; these differ in their composition and mode of synthesis of the polyketide synthesized. These two major types of PKS enzymes are commonly referred to as Type I or "modular" and Type II "iterative" PKS enzymes.

In the Type I or modular PKS enzyme group, a set of separate catalytic active sites (each active site is termed a "domain", and a set thereof is termed a "module") exists for each cycle of carbon chain elongation and modification in the polyketide synthesis pathway. The typical modular PKS is composed of several large polypeptides, which can be segregated from amino to carboxy termini into a loading module, multiple extender modules, and a releasing (or thioesterase) domain. The PKS enzyme known as 6-deoxyerythronolide B synthase (DEBS) is a Type I PKS. In DEBS, there is a loading module, six extender modules, and a thioesterase (TE) domain. The loading module, six extender modules, and TE of DEBS are present on three separate proteins (designated DEBS-1, DEBS-2, and DEBS-3, with two extender modules per protein). Each of the DEBS polypeptides is encoded by a separate open reading frame (ORF) or gene; these genes are known as *eryA1*, *eryAII*, and *eryAIII*. See Caffrey *et al.*, 1992, *FEBS Letters* 304: 205, and U.S. Patent No. 5,824,513, each of which is incorporated herein by reference.

Generally, the loading module is responsible for binding the first building block used to synthesize the polyketide and transferring it to the first extender module. The loading module of DEBS consists of an acyltransferase (AT) domain and an acyl carrier protein (ACP) domain. Another type of loading module utilizes an inactivated ketosynthase (KS) domain and AT and ACP domains. This inactivated KS is in some instances called KS^Q, where the superscript letter is the abbreviation for the amino acid, glutamine, that is

present instead of the active site cysteine required for ketosynthase activity. In other PKS enzymes, including the FK-506 PKS, the loading module incorporates an unusual starter unit and is composed of a CoA ligase like activity domain. In any event, the loading module recognizes a particular acyl-CoA (usually acetyl or propionyl but sometimes butyryl or 5 other acyl-CoA) and transfers it as a thiol ester to the ACP of the loading module.

The AT on each of the extender modules recognizes a particular extender-CoA (malonyl or alpha-substituted malonyl, i.e., methylmalonyl, ethylmalonyl, and 2-hydroxymalonyl) and transfers it to the ACP of that extender module to form a thioester. Each extender module is responsible for accepting a compound from a prior module, 10 binding a building block, attaching the building block to the compound from the prior module, optionally performing one or more additional functions, and transferring the resulting compound to the next module.

Each extender module of a modular PKS contains a KS, AT, ACP, and zero, one, two, or three domains that modify the beta-carbon of the growing polyketide chain. A 15 typical (non-loading) minimal Type I PKS extender module is exemplified by extender module three of DEBS, which contains a KS domain, an AT domain, and an ACP domain. These three domains are sufficient to activate a 2-carbon extender unit and attach it to the growing polyketide molecule. The next extender module, in turn, is responsible for attaching the next building block and transferring the growing compound to the next 20 extender module until synthesis is complete.

Once the PKS is primed with acyl- and malonyl-ACPs, the acyl group of the loading module is transferred to form a thiol ester (trans-esterification) at the KS of the first extender module; at this stage, extender module one possesses an acyl-KS and a malonyl (or substituted malonyl) ACP. The acyl group derived from the loading module is then 25 covalently attached to the alpha-carbon of the malonyl group to form a carbon-carbon bond, driven by concomitant decarboxylation, and generating a new acyl-ACP that has a backbone two carbons longer than the loading building block (elongation or extension).

The polyketide chain, growing by two carbons each extender module, is sequentially passed as covalently bound thiol esters from extender module to extender module, in an 30 assembly line-like process. The carbon chain produced by this process alone would possess a ketone at every other carbon atom, producing a polyketone, from which the name polyketide arises. Most commonly, however, additional enzymatic activities modify the beta

keto group of each two carbon unit just after it has been added to the growing polyketide chain but before it is transferred to the next module.

Thus, in addition to the minimal module containing KS, AT, and ACP domains necessary to form the carbon-carbon bond, and as noted above, other domains that modify 5 the beta-carbonyl moiety can be present. Thus, modules may contain a ketoreductase (KR) domain that reduces the keto group to an alcohol. Modules may also contain a KR domain plus a dehydratase (DH) domain that dehydrates the alcohol to a double bond. Modules may also contain a KR domain, a DH domain, and an enoylreductase (ER) domain that converts 10 the double bond product to a saturated single bond using the beta carbon as a methylene function. An extender module can also contain other enzymatic activities, such as, for example, a methylase or dimethylase activity.

After traversing the final extender module, the polyketide encounters a releasing domain that cleaves the polyketide from the PKS and typically cyclizes the polyketide. For example, final synthesis of 6-dEB is regulated by a TE domain located at the end of 15 extender module six. In the synthesis of 6-dEB, the TE domain catalyzes cyclization of the macrolide ring by formation of an ester linkage. In FK-506, FK-520, rapamycin, and similar polyketides, the TE activity is replaced by a RapP (for rapamycin) or RapP like activity that makes a linkage incorporating a pipecolate acid residue. The enzymatic activity that catalyzes this incorporation for the rapamycin enzyme is known as RapP, encoded by the 20 *rapP* gene. The polyketide can be modified further by tailoring enzymes; these enzymes add carbohydrate groups or methyl groups, or make other modifications, i.e., oxidation or reduction, on the polyketide core molecule. For example, 6-dEB is hydroxylated at C-6 and C-12 and glycosylated at C-3 and C-5 in the synthesis of erythromycin A.

In Type I PKS polypeptides, the order of catalytic domains is conserved. When all 25 beta-keto processing domains are present in a module, the order of domains in that module from N-to-C-terminus is always KS, AT, DH, ER, KR, and ACP. Some or all of the beta-keto processing domains may be missing in particular modules, but the order of the domains present in a module remains the same. The order of domains within modules is believed to be important for proper folding of the PKS polypeptides into an active complex. Importantly, 30 there is considerable flexibility in PKS enzymes, which allows for the genetic engineering of novel catalytic complexes. The engineering of these enzymes is achieved by modifying, adding, or deleting domains, or replacing them with those taken from other Type I PKS enzymes. It is also achieved by deleting, replacing, or adding entire modules with those

taken from other sources. A genetically engineered PKS complex should of course have the ability to catalyze the synthesis of the product predicted from the genetic alterations made.

Alignments of the many available amino acid sequences for Type I PKS enzymes has approximately defined the boundaries of the various catalytic domains. Sequence 5 alignments also have revealed linker regions between the catalytic domains and at the N- and C-termini of individual polypeptides. The sequences of these linker regions are less well conserved than are those for the catalytic domains, which is in part how linker regions are identified. Linker regions can be important for proper association between domains and between the individual polypeptides that comprise the PKS complex. One can thus view the 10 linkers and domains together as creating a scaffold on which the domains and modules are positioned in the correct orientation to be active. This organization and positioning, if retained, permits PKS domains of different or identical substrate specificities to be substituted (usually at the DNA level) between PKS enzymes by various available methodologies. In selecting the boundaries of, for example, an AT replacement, one can 15 thus make the replacement so as to retain the linkers of the recipient PKS or to replace them with the linkers of the donor PKS AT domain, or, preferably, make both constructs to ensure that the correct linker regions between the KS and AT domains have been included in at least one of the engineered enzymes. Thus, there is considerable flexibility in the 20 design of new PKS enzymes with the result that known polyketides can be produced more effectively, and novel polyketides useful as pharmaceuticals or for other purposes can be made.

By appropriate application of recombinant DNA technology, a wide variety of polyketides can be prepared in a variety of different host cells provided one has access to nucleic acid compounds that encode PKS proteins and polyketide modification enzymes. 25 The present invention helps meet the need for such nucleic acid compounds by providing recombinant vectors that encode the FK-520 PKS enzyme and various FK-520 modification enzymes. Moreover, while the FK-506 and FK-520 polyketides have many useful activities, there remains a need for compounds with similar useful activities but with better pharmacokinetic profile and metabolism and fewer side-effects. The present invention helps 30 meet the need for such compounds as well.

Summary of the Invention

SUBSTITUTE SHEET (RULE 26)

In one embodiment, the present invention provides recombinant DNA vectors that encode all or part of the FK-520 PKS enzyme. Illustrative vectors of the invention include cosmid pKOS034-120, pKOS034-124, pKOS065-C31, pKOS065-C3, pKOS065-M27, and pKOS065-M21. The invention also provides nucleic acid compounds that encode the various domains of the FK-520 PKS, i.e., the KS, AT, ACP, KR, DH, and ER domains. These compounds can be readily used, alone or in combination with nucleic acids encoding other FK-520 or non-FK-520 PKS domains, as intermediates in the construction of recombinant vectors that encode all or part of PKS enzymes that make novel polyketides.

5 The invention also provides isolated nucleic acids that encode all or part of one or more modules of the FK-520 PKS, each module comprising a ketosynthase activity, an acyl transferase activity, and an acyl carrier protein activity. The invention provides an isolated nucleic acid that encodes one or more open reading frames of FK-520 PKS genes, said open reading frames comprising coding sequences for a CoA ligase activity, an NRPS activity, or 10 two or more extender modules. The invention also provides recombinant expression vectors 15 containing these nucleic acids.

In another embodiment, the invention provides isolated nucleic acids that encode all or a part of a PKS that contains at least one module in which at least one of the domains in the module is a domain from a non-FK-520 PKS and at least one domain is from the FK-520 PKS. The non-FK-520 PKS domain or module originates from the rapamycin PKS, the 20 FK-506 PKS, DEBS, or another PKS. The invention also provides recombinant expression vectors containing these nucleic acids.

25 In another embodiment, the invention provides a method of preparing a polyketide. said method comprising transforming a host cell with a recombinant DNA vector that encodes at least one module of a PKS, said module comprising at least one FK-520 PKS domain, and culturing said host cell under conditions such that said PKS is produced and catalyzes synthesis of said polyketide. In one aspect, the method is practiced with a *Streptomyces* host cell. In another aspect, the polyketide produced is FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-520. In another aspect, the polyketide produced is a polyketide related in structure to FK-506 or rapamycin.

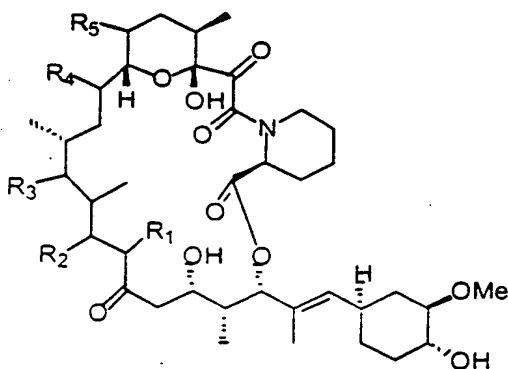
30 In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of ethylmalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require ethylmalonyl CoA for biosynthesis.

The invention also provides recombinant nucleic acids that encode AT domains specific for ethylmalonyl CoA. Thus, the compounds of the invention can be used to produce polyketides requiring ethylmalonyl CoA in host cells that otherwise are unable to produce such polyketides.

5 In another embodiment, the invention provides a set of genes in recombinant form sufficient for the synthesis of 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA in a heterologous host cell. These genes and the methods of the invention enable one to create recombinant host cells with the ability to produce polyketides or other compounds that require 2-hydroxymalonyl CoA for biosynthesis. The invention also provides recombinant 10 nucleic acids that encode AT domains specific for 2-hydroxymalonyl CoA and 2-methoxymalonyl CoA. Thus, the compounds of the invention can be used to produce polyketides requiring 2-hydroxymalonyl CoA or 2-methoxymalonyl CoA in host cells that are otherwise unable to produce such polyketides.

15 In another embodiment, the invention provides a compound related in structure to FK-520 or FK-506 that is useful in the treatment of a medical condition. These compounds include compounds in which the C-13 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. Such compounds are less susceptible to the main *in vivo* pathway of degradation for FK-520 and FK-506 and related compounds and thus exhibit an improved pharmacokinetic profile. The compounds of the 20 invention also include compounds in which the C-15 methoxy group is replaced by a moiety selected from the group consisting of hydrogen, methyl, and ethyl moieties. The compounds of the invention also include the above compounds further modified by chemical methodology to produce derivatives such as, but not limited to, the C-18 hydroxyl derivatives, which have potent neurotrophin but not immunosuppression activities.

25 Thus, the invention provides polyketides having the structure:



wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen or hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, 5 methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506. The invention provides these compounds in purified form and in pharmaceutical compositions.

In another embodiment, the invention provides a method for treating a medical condition by administering a pharmaceutically efficacious dose of a compound of the 10 invention. The compounds of the invention may be administered to achieve immunosuppression or to stimulate nerve growth and regeneration.

These and other embodiments and aspects of the invention will be more fully understood after consideration of the attached Drawings and their brief description below, together with the detailed description, examples, and claims that follow.

15

Brief Description of the Drawings

Figure 1 shows a diagram of the FK-520 biosynthetic gene cluster. The top line provides a scale in kilobase pairs (kb). The second line shows a restriction map with selected restriction enzyme recognition sequences indicated. K is *Kpn*I; X is *Xho*I, S is *Sac*I; 20 P is *Pst*I; and E is *Eco*RI. The third line indicates the position of FK-520 PKS and related genes. Genes are abbreviated with a one letter designation, i.e., C is *fkb*C. Immediately under the third line are numbered segments showing where the loading module (L) and ten different extender modules (numbered 1 - 10) are encoded on the various genes shown. At the bottom of the Figure, the DNA inserts of various cosmids of the invention (i.e., 34-124 25 is cosmid pKOS034-124) are shown in alignment with the FK-520 biosynthetic gene cluster.

Figure 2 shows the loading module (load), the ten extender modules, and the peptide synthetase domain of the FK-520 PKS, together with, on the top line, the genes that encode the various domains and modules. Also shown are the various intermediates in FK-520 30 biosynthesis, as well as the structure of FK-520, with carbons 13, 15, 21, and 31 numbered. The various domains of each module and subdomains of the loading module are also shown. The darkened circles showing the DH domains in modules 2, 3, and 4 indicate that the dehydratase domain is not functional as a dehydratase; this domain may affect the

stereochemistry at the corresponding position in the polyketide. The substituents on the FK-520 structure that result from the action of non-PKS enzymes are also indicated by arrows, together with the types of enzymes or the genes that code for the enzymes that mediate the action. Although the methyltransferase is shown acting at the C-13 and C-15 hydroxyl groups after release of the polyketide from the PKS, the methyltransferase may act on the 2-hydroxymalonyl substrate prior to or contemporaneously with its incorporation during polyketide synthesis.

Figure 3 shows a close-up view of the left end of the FK-520 gene cluster, which contains at least ten additional genes. The ethyl side chain on carbon 21 of FK-520 (Figure 2) is derived from an ethylmalonyl CoA extender unit that is incorporated by an ethylmalonyl specific AT domain in extender module 4 of the PKS. At least four of the genes in this region code for enzymes involved in ethylmalonyl biosynthesis. The polyhydroxybutyrate depolymerase is involved in maintaining hydroxybutyryl-CoA pools during FK-520 production. Polyhydroxybutyrate accumulates during vegetative growth and disappears during stationary phase in other *Streptomyces* (Ranade and Vining, 1993, *Can. J. Microbiol.* 39:377). Open reading frames with unknown function are indicated with a question mark.

Figure 4 shows a biosynthetic pathway for the biosynthesis of ethylmalonyl CoA from acetoacetyl CoA consistent with the function assigned to four of the genes in the FK-520 gene cluster shown in Figure 3.

Figure 5 shows a close-up view of the right-end of the FK-520 PKS gene cluster (and of the sequences on cosmid pKOS065-C31). The genes shown include *fkbD*, *fkbM* (a methyl transferase that methylates the hydroxyl group on C-31 of FK-520), *fkbN* (a homolog of a gene described as a regulator of cholesterol oxidase and that is believed to be a transcriptional activator), *fkbQ* (a type II thioesterase, which can increase polyketide production levels), and *fkbS* (a crotonyl-CoA reductase involved in the biosynthesis of ethylmalonyl CoA).

Figure 6 shows the proposed degradative pathway for tacrolimus (FK-506) metabolism.

Figure 7 shows a schematic process for the construction of recombinant PKS genes of the invention that encode PKS enzymes that produce 13-desmethoxy FK-506 and FK-520 polyketides of the invention, as described in Example 4, below.

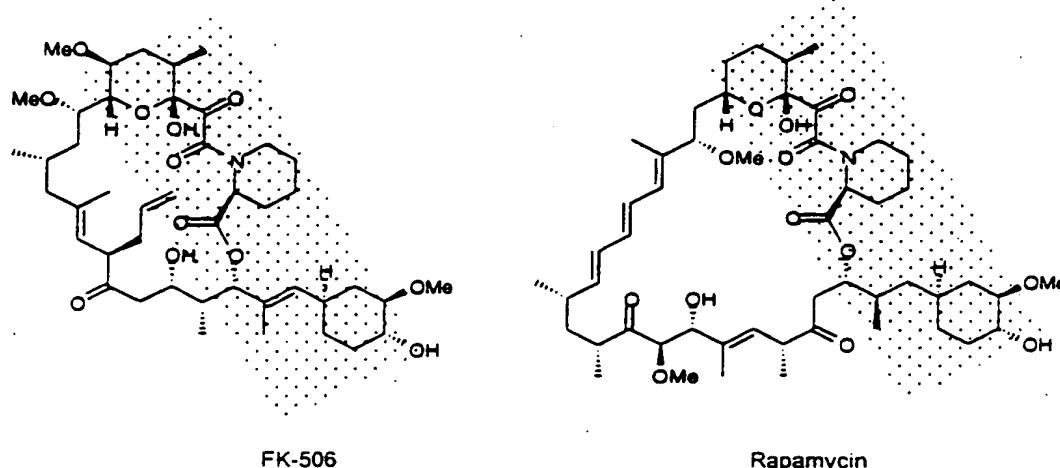
Figure 8, in Parts A and B, shows certain compounds of the invention preferred for dermal application in Part A and a synthetic route for making those compounds in Part B.

Detailed Description of the Invention

- 5 Given the valuable pharmaceutical properties of polyketides, there is a need for methods and reagents for producing large quantities of polyketides, as well as for producing related compounds not found in nature. The present invention provides such methods and reagents, with particular application to methods and reagents for producing the polyketides known as FK-520, also known as ascomycin or L-683,590 (see Holt *et al.*, 1993, *JACS* 115:9925), and FK-506, also known as tacrolimus. Tacrolimus is a macrolide immunosuppressant used to prevent or treat rejection of transplanted heart, kidney, liver, lung, pancreas, and small bowel allografts. The drug is also useful for the prevention and treatment of graft-versus-host disease in patients receiving bone marrow transplants, and for the treatment of severe, refractory uveitis. There have been additional reports of the 10 unapproved use of tacrolimus for other conditions, including alopecia universalis, autoimmune chronic active hepatitis, inflammatory bowel disease, multiple sclerosis, primary biliary cirrhosis, and scleroderma. The invention provides methods and reagents for making novel polyketides related in structure to FK-520 and FK-506, and structurally related polyketides such as rapamycin.

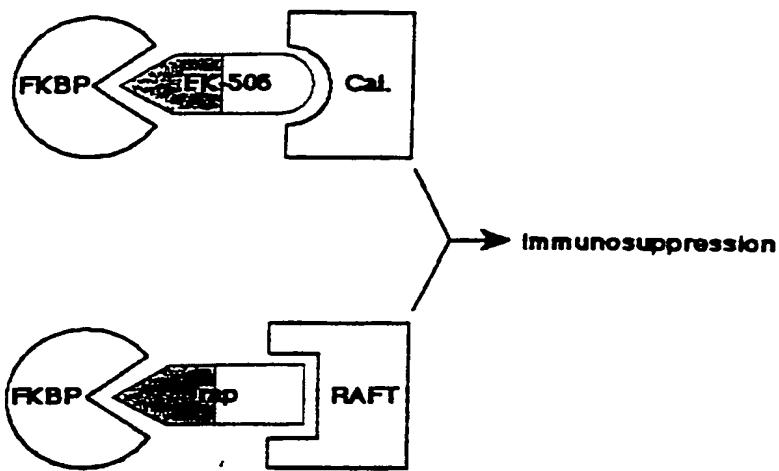
15

20 The FK-506 and rapamycin polyketides are potent immunosuppressants, with chemical structures shown below.



FK-520 differs from FK-506 in that it lacks the allyl group at C-21 of FK-506, having instead an ethyl group at that position, and has similar activity to FK-506, albeit reduced immunosuppressive activity.

These compounds act through initial formation of an intermediate complex with 5 protein "immunophilins" known as FKBP (FK-506 binding proteins), including FKBP-12. Immunophilins are a class of cytosolic proteins that form complexes with molecules such as FK-506, FK-520, and rapamycin that in turn serve as ligands for other cellular targets involved in signal transduction. Binding of FK-506, FK-520, and rapamycin to FKBP occurs through the structurally similar segments of the polyketide molecules, known as the 10 "FKBP-binding domain" (as generally but not precisely indicated by the stippled regions in the structures above). The FK-506-FKBP complex then binds calcineurin, while the rapamycin-FKBP complex binds to a protein known as RAFT-1. Binding of the FKBP-polyketide complex to these second proteins occurs through the dissimilar regions of the drugs known as the "effector" domains.

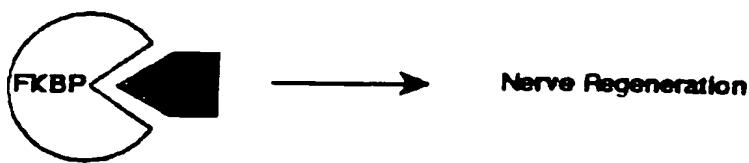


15 The three component FKBP-polyketide-effector complex is required for signal transduction and subsequent immunosuppressive activity of FK-506, FK-520, and rapamycin. Modifications in the effector domains of FK-506, FK-520, and rapamycin that destroy binding to the effector proteins (calcineurin or RAFT) lead to loss of 20 immunosuppressive activity, even though FKBP binding is unaffected. Further, such analogs antagonize the immunosuppressive effects of the parent polyketides, because they compete for FKBP. Such non-immunosuppressive analogs also show reduced toxicity (see Dumont *et al.*, 1992, *Journal of Experimental Medicine* 176, 751-760), indicating that much of the toxicity of these drugs is not linked to FKBP binding.

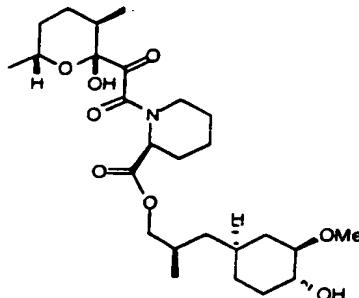
In addition to immunosuppressive activity, FK-520, FK-506, and rapamycin have neurotrophic activity. In the central nervous system and in peripheral nerves, immunophilins are referred to as 'neuroimmunophilins'. The neuroimmunophilin FKBP is markedly enriched in the central nervous system and in peripheral nerves. Molecules that bind to the 5 neuroimmunophilin FKBP, such as FK-506 and FK-520, have the remarkable effect of stimulating nerve growth. *In vitro*, they act as neurotrophins, i.e., they promote neurite outgrowth in NGF-treated PC12 cells and in sensory neuronal cultures, and in intact animals, they promote regrowth of damaged facial and sciatic nerves, and repair lesioned serotonin and dopamine neurons in the brain. See Gold *et al.*, Jun. 1999, *J. Pharm. Exp. Ther.* 289(3): 1202-1210; Lyons *et al.*, 1994, *Proc. National Academy of Science* 91: 3191-3195; Gold *et al.*, 1995, *Journal of Neuroscience* 15: 7509-7516; and Steiner *et al.*, 1997, *Proc. National Academy of Science* 94: 2019-2024. Further, the restored central and peripheral neurons appear to be functional.

Compared to protein neurotrophic molecules (BNDF, NGF, etc.), the small-molecule neurotrophins such as FK-506, FK-520, and rapamycin have different, and often advantageous, properties. First, whereas protein neurotrophins are difficult to deliver to their intended site of action and may require intra-cranial injection, the small-molecule neurotrophins display excellent bioavailability; they are active when administered subcutaneously and orally. Second, whereas protein neurotrophins show quite specific effects, the small-molecule neurotrophins show rather broad effects. Finally, whereas protein neurotrophins often show effects on normal sensory nerves, the small-molecule neurotrophins do not induce aberrant sprouting of normal neuronal processes and seem to affect damaged nerves specifically. Neuroimmunophilin ligands have potential therapeutic utility in a variety of disorders involving nerve degeneration (e.g. multiple sclerosis, 20 Parkinson's disease, Alzheimer's disease, stroke, traumatic spinal cord and brain injury, peripheral neuropathies).

Recent studies have shown that the immunosuppressive and neurite outgrowth activity of FK-506, FK-520, and rapamycin can be separated; the neuroregenerative activity in the absence of immunosuppressive activity is retained by agents which bind to FKBP but 25 not to the effector proteins calcineurin or RAFT. See Steiner *et al.*, 1997, *Nature Medicine* 3: 421-428.



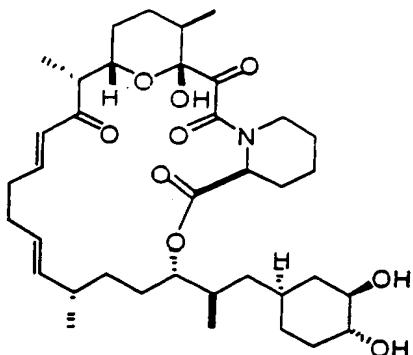
Available structure-activity data show that the important features for neurotrophic activity of rapamycin, FK-520, and FK-506 lie within the common, contiguous segments of the macrolide ring that bind to FKBP. This portion of the molecule is termed the "FKBP binding domain" (see VanDuyne *et al.*, 1993, *Journal of Molecular Biology* 229: 105-124.). Nevertheless, the effector domains of the parent macrolides contribute to conformational rigidity of the binding domain and thus indirectly contribute to FKBP binding.



"FKBP binding domain"

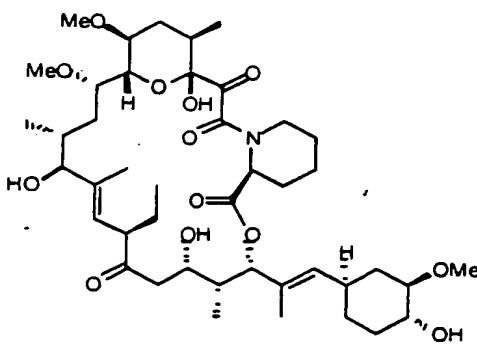
There are a number of other reported analogs of FK-506, FK-520, and rapamycin that bind to FKBP but not the effector protein calcineurin or RAFT. These analogs show effects on nerve regeneration without immunosuppressive effects.

Naturally occurring FK-520 and FK-506 analogs include the antascomycins, which are FK-506-like macrolides that lack the functional groups of FK-506 that bind to calcineurin (see Fehr *et al.*, 1996, *The Journal of Antibiotics* 49: 230-233). These molecules bind FKBP as effectively as does FK-506; they antagonize the effects of both FK-506 and rapamycin, yet lack immunosuppressive activity.

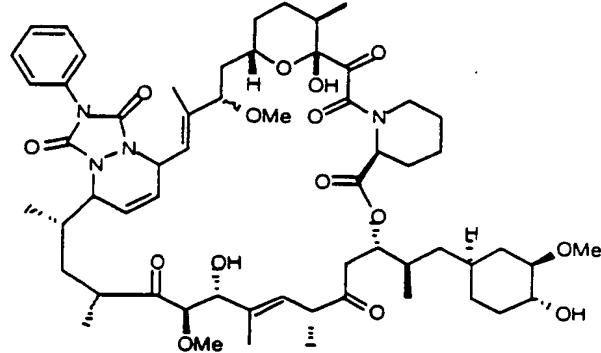


Antascomycin A

Other analogs can be produced by chemically modifying FK-506, FK-520, or rapamycin. One approach to obtaining neuroimmunophilin ligands is to destroy the effector binding region of FK-506, FK-520, or rapamycin by chemical modification. While the 5 chemical modifications permitted on the parent compounds are quite limited, some useful chemically modified analogs exist. The FK-520 analog L-685,818 ($ED_{50} = 0.7$ nM for FKBP binding; see Dumont *et al.*, 1992), and the rapamycin analog WAY-124,466 ($IC_{50} = 12.5$ nM; see Ocain *et al.*, 1993, *Biochemistry Biophysical Research Communications* 192: 1340-134693) are about as effective as FK-506, FK-520, and rapamycin at promoting 10 neurite outgrowth in sensory neurons (see Steiner *et al.*, 1997).



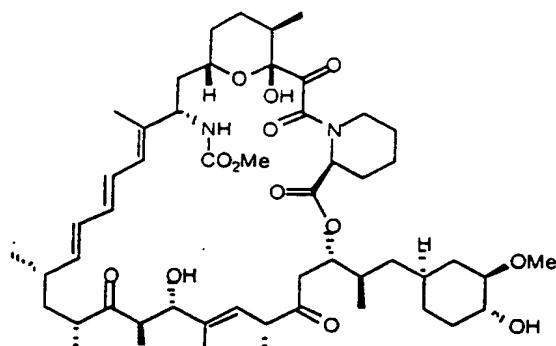
L-685,818



WAY-124,466

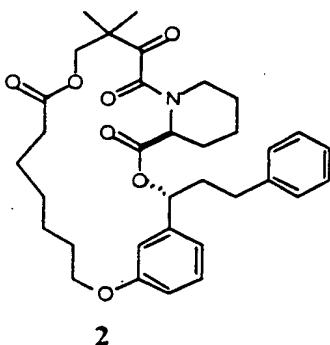
One of the few positions of rapamycin that is readily amenable to chemical 15 modification is the allylic 16-methoxy group; this reactive group is readily exchanged by acid-catalyzed nucleophilic substitution. Replacement of the 16-methoxy group of rapamycin with a variety of bulky groups has produced analogs showing selective loss of immunosuppressive activity while retaining FKBP-binding (see Luengo *et al.*, 1995, *Chemistry & Biology* 2: 471-481). One of the best compounds, 1, below, shows complete

loss of activity in the splenocyte proliferation assay with only a 10-fold reduction in binding to FKBP.

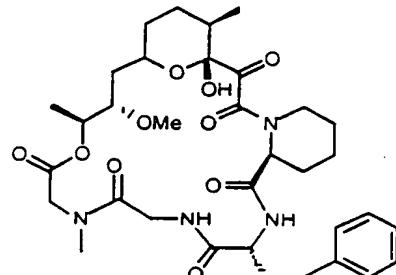


1

5 There are also synthetic analogs of FKBP binding domains. These compounds reflect an approach to obtaining neuroimmunophilin ligands based on "rationally designed" molecules that retain the FKBP-binding region in an appropriate conformation for binding to FKBP, but do not possess the effector binding regions. In one example, the ends of the FKBP binding domain were tethered by hydrocarbon chains (see Holt *et al.*, 1993, *Journal*
10 *of the American Chemical Society* 115: 9925-9938); the best analog, 2, below, binds to FKBP about as well as FK-506. In a similar approach, the ends of the FKBP binding domain were tethered by a tripeptide to give analog 3, below, which binds to FKBP about 20-fold poorer than FK-506. These compounds are anticipated to have neuroimmunophilin binding activity.



2



3

15

In a primate MPTP model of Parkinson's disease, administration of FKBP ligand
15 GPI-1046 caused brain cells to regenerate and behavioral measures to improve. MPTP is a
neurotoxin, which, when administered to animals, selectively damages nigral-striatal
20 dopamine neurons in the brain, mimicking the damage caused by Parkinson's disease.
Whereas, before treatment, animals were unable to use affected limbs, the FKBP ligand

restored the ability of animals to feed themselves and gave improvements in measures of locomotor activity, neurological outcome, and fine motor control. There were also corresponding increases in regrowth of damaged nerve terminals. These results demonstrate the utility of FKBP ligands for treatment of diseases of the CNS.

5 From the above description, two general approaches towards the design of non-immunosuppressant, neuroimmunophilin ligands can be seen. The first involves the construction of constrained cyclic analogs of FK-506 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP. The advantages of this approach are that the conformation of the analogs can be accurately modeled and predicted by
10 computational methods, and the analogs closely resemble parent molecules that have proven pharmacological properties. A disadvantage is that the difficult chemistry limits the numbers and types of compounds that can be prepared. The second approach involves the trial and error construction of acyclic analogs of the FKBP binding domain by conventional medicinal chemistry. The advantages to this approach are that the chemistry is suitable for
15 production of the numerous compounds needed for such interactive chemistry-bioassay approaches. The disadvantages are that the molecular types of compounds that have emerged have no known history of appropriate pharmacological properties, have rather labile ester functional groups, and are too conformationally mobile to allow accurate prediction of conformational properties.

20 The present invention provides useful methods and reagents related to the first approach, but with significant advantages. The invention provides recombinant PKS genes that produce a wide variety of polyketides that cannot otherwise be readily synthesized by chemical methodology alone. Moreover, the present invention provides polyketides that have either or both of the desired immunosuppressive and neurotrophic activities, some of
25 which are produced only by fermentation and others of which are produced by fermentation and chemical modification. Thus, in one aspect, the invention provides compounds that optimally bind to FKBP but do not bind to the effector proteins. The methods and reagents of the invention can be used to prepare numerous constrained cyclic analogs of FK-520 in which the FKBP binding domain is fixed in a conformation optimal for binding to FKBP.
30 Such compounds will show neuroimmunophilin binding (neurotrophic) but not immunosuppressive effects. The invention also allows direct manipulation of FK-520 and related chemical structures via genetic engineering of the enzymes involved in the biosynthesis of FK-520 (as well as related compounds, such as FK-506 and rapamycin);

similar chemical modifications are simply not possible because of the complexity of the structures. The invention can also be used to introduce "chemical handles" into normally inert positions that permit subsequent chemical modifications.

Several general approaches to achieve the development of novel neuroimmunophilin 5 ligands are facilitated by the methods and reagents of the present invention. One approach is to make "point mutations" of the functional groups of the parent FK-520 structure that bind to the effector molecules to eliminate their binding potential. These types of structural modifications are difficult to perform by chemical modification, but can be readily accomplished with the methods and reagents of the invention.

10 A second, more extensive approach facilitated by the present invention is to utilize molecular modeling to predict optimal structures *ab initio* that bind to FKBP but not effector molecules. Using the available X-ray crystal structure of FK-520 (or FK-506) bound to FKBP, molecular modeling can be used to predict polyketides that should optimally bind to FKBP but not calcineurin. Various macrolide structures can be generated 15 by linking the ends of the FKBP-binding domain with "all possible" polyketide chains of variable length and substitution patterns that can be prepared by genetic manipulation of the FK-520 or FK-506 PKS gene cluster in accordance with the methods of the invention. The ground state conformations of the virtual library can be determined, and compounds that possess binding domains most likely to bind well to FKBP can be prepared and tested.

20 Once a compound is identified in accordance with the above approaches, the invention can be used to generate a focused library of analogs around the lead candidate, to "fine tune" the compound for optimal properties. Finally, the genetic engineering methods of the invention can be directed towards producing "chemical handles" that enable medicinal chemists to modify positions of the molecule previously inert to chemical 25 modification. This opens the path to previously prohibited chemical optimization of lead compounds by time-proven approaches.

Moreover, the present invention provides polyketide compounds and the recombinant genes for the PKS enzymes that produce the compounds that have significant advantages over FK-506 and FK-520 and their analogs. The metabolism and 30 pharmacokinetics of tacrolimus has been extensively studied, and FK-520 is believed to be similar in these respects. Absorption of tacrolimus is rapid, variable, and incomplete from the gastrointestinal tract (Harrison's Principles of Internal Medicine, 14th edition, 1998, McGraw Hill, 14, 20, 21, 64-67). The mean bioavailability of the oral dosage form is 27%.

(range 5 to 65%). The volume of distribution (V_oD) based on plasma is 5 to 65 L per kg of body weight (L/kg), and is much higher than the V_oD based on whole blood concentrations, the difference reflecting the binding of tacrolimus to red blood cells. Whole blood concentrations may be 12 to 67 times the plasma concentrations. Protein binding is 5 high (75 to 99%), primarily to albumin and alpha₁-acid glycoprotein. The half-life for distribution is 0.9 hour; elimination is biphasic and variable: terminal-11.3 hr (range, 3.5 to 40.5 hours). The time to peak concentration is 0.5 to 4 hours after oral administration.

10 Tacrolimus is metabolized primarily by cytochrome P450 3A enzymes in the liver and small intestine. The drug is extensively metabolized with less than 1% excreted unchanged in urine. Because hepatic dysfunction decreases clearance of tacrolimus, doses have to be reduced substantially in primary graft non-function, especially in children. In addition, drugs that induce the cytochrome P450 3A enzymes reduce tacrolimus levels, while drugs that inhibit these P450s increase tacrolimus levels. Tacrolimus bioavailability doubles with co-administration of ketoconazole, a drug that inhibits P450 3A. See, Vincent 15 *et al.*, 1992, *In vitro metabolism of FK-506 in rat, rabbit, and human liver microsomes: Identification of a major metabolite and of cytochrome P450 3A as the major enzymes responsible for its metabolism*, *Arch. Biochem. Biophys.* 294: 454-460; Iwasaki *et al.*, 1993, *Isolation, identification, and biological activities of oxidative metabolites of FK-506, a potent immunosuppressive macrolide lactone*, *Drug Metabolism & Disposition* 21: 971-977; 20 Shiraga *et al.*, 1994, *Metabolism of FK-506, a potent immunosuppressive agent, by cytochrome P450 3A enzymes in rat, dog, and human liver microsomes*, *Biochem. Pharmacol.* 47: 727-735; and Iwasaki *et al.*, 1995, *Further metabolism of FK-506 (Tacrolimus); Identification and biological activities of the metabolites oxidized at multiple sites of FK-506*, *Drug Metabolism & Disposition* 23: 28-34. The cytochrome P450 3A 25 subfamily of isozymes has been implicated as important in this degradative process.

Structures of the eight isolated metabolites formed by liver microsomes are shown in Figure 6. Four metabolites of FK-506 involve demethylation of the oxygens on carbons 13, 15, and 31, and hydroxylation of carbon 12. The 13-demethylated (hydroxy) compounds undergo cyclizations of the 13-hydroxy at C-10 to give M-I, M-VI and M-VII, and the 12-hydroxy metabolite at C-10 to give I. Another four metabolites formed by oxidation of the four metabolites mentioned above were isolated by liver microsomes from dexamethasone treated rats. Three of these are metabolites doubly demethylated at the methoxy groups on carbons 15 and 31 (M-V), 13 and 31 (M-VI), and 13 and 15 (M-VII). The fourth, M-VIII,

was the metabolite produced after demethylation of the 31-methoxy group, followed by formation of a fused ring system by further oxidation. Among the eight metabolites, M-II has immunosuppressive activity comparable to that of FK-506, whereas the other metabolites exhibit weak or negligible activities. Importantly, the major metabolite of 5 human, dog, and rat liver microsomes is the 13-demethylated and cyclized FK-506 (M-I).

Thus, the major metabolism of FK-506 proceeds via 13-demethylation followed by cyclization to the inactive M-I, this representing about 90% of the metabolic products after a 10 minute incubation with liver microsomes. Analogs of tacrolimus that do not possess a C-13 methoxy group would not be susceptible to the first and most important 10 biotransformation in the destructive metabolism of tacrolimus (i.e. cyclization of 13-hydroxy to C-10). Thus, a 13-desmethoxy analog of FK-506 should have a longer half-life in the body than does FK-506. The C-13 methoxy group is believed not to be required for binding to FKBP or calcineurin. The C-13 methoxy is not present on the identical position 15 of rapamycin, which binds to FKBP with equipotent affinity as tacrolimus. Also, analysis of the 3-dimensional structure of the FKBP-tacrolimus-calcineurin complex shows that the C-13 methoxy has no interaction with FKBP and only a minor interaction with calcineurin. The present invention provides C-13-desmethoxy analogs of FK-506 and FK-520, as well as the recombinant genes that encode the PKS enzymes that catalyze their synthesis and host cells that produce the compounds.

20 These compounds exhibit, relative to their naturally occurring counterparts, prolonged immunosuppressive action *in vivo*, thereby allowing a lower dosage and/or reduced frequency of administration. Dosing is more predictable, because the variability in FK-506 dosage is largely due to variation of metabolism rate. FK-506 levels in blood can vary widely depending on interactions with drugs that induce or inhibit cytochrome P450 25 3A (summarized in USP Drug Information for the Health Care Professional). Of particular importance are the numerous drugs that inhibit or compete for CYP 3A, because they increase FK-506 blood levels and lead to toxicity (Prograf package insert, Fujisawa US, Rev 4/97, Rec 6/97). Also important are the drugs that induce P450 3A (e.g. Dexamethasone), because they decrease FK-506 blood levels and reduce efficacy. Because 30 the major site of CYP 3A action on FK-506 is removed in the analogs provided by the present invention, those analogs are not as susceptible to drug interactions as the naturally occurring compounds.

Hyperglycemia, nephrotoxicity, and neurotoxicity are the most significant adverse effects resulting from the use of FK-506 and are believed to be similar for FK-520. Because these effects appear to occur primarily by the same mechanism as the immunosuppressive action (i.e. FKBP-calcineurin interaction), the intrinsic toxicity of the desmethoxy analogs 5 may be similar to FK-506. However, toxicity of FK-506 is dose related and correlates with high blood levels of the drug (Prograf package insert, Fujisawa[®] US, Rev 4/97, Rec 6/97). Because the levels of the compounds provided by the present invention should be more controllable, the incidence of toxicity should be significantly decreased with the 13-desmethoxy analogs. Some reports show that certain FK-506 metabolites are more toxic 10 than FK-506 itself, and this provides an additional reason to expect that a CYP 3A resistant analog can have lower toxicity and a higher therapeutic index.

Thus, the present invention provides novel compounds related in structure to FK-506 and FK-520 but with improved properties. The invention also provides methods for making these compounds by fermentation of recombinant host cells, as well as the 15 recombinant host cells, the recombinant vectors in those host cells, and the recombinant proteins encoded by those vectors. The present invention also provides other valuable materials useful in the construction of these recombinant vectors that have many other important applications as well. In particular, the present invention provides the FK-520 PKS genes, as well as certain genes involved in the biosynthesis of FK-520 in recombinant form.

20 FK-520 is produced at relatively low levels in the naturally occurring cells, *Streptomyces hygroscopicus* var. *ascomyceticus*, in which it was first identified. Thus, another benefit provided by the recombinant FK-520 PKS and related genes of the present invention is the ability to produce FK-520 in greater quantities in the recombinant host cells provided by the invention. The invention also provides methods for making novel FK-520 25 analogs, in addition to the desmethoxy analogs described above, and derivatives in recombinant host cells of any origin.

The biosynthesis of FK-520 involves the action of several enzymes. The FK-520 PKS enzyme, which is composed of the *fkbA*, *fkbB*, *fkbC*, and *fkbP* gene products, synthesizes the core structure of the molecule. There is also a hydroxylation at C-9 mediated 30 by the P450 hydroxylase that is the *fkbD* gene product and that is oxidized by the *fkbO* gene product to result in the formation of a keto group at C-9. There is also a methylation at C-31 that is mediated by an O-methyltransferase that is the *fkbM* gene product. There are also methylations at the C-13 and C-15 positions by a methyltransferase believed to be encoded

by the *fkbG* gene; this methyltransferase may act on the hydroxymalonyl CoA substrates prior to binding of the substrate to the AT domains of the PKS during polyketide synthesis. The present invention provides the genes encoding these enzymes in recombinant form. The invention also provides the genes encoding the enzymes involved in ethylmalonyl CoA and 5 2-hydroxymalonyl CoA biosynthesis in recombinant form. Moreover, the invention provides *Streptomyces hygroscopicus* var. *ascomyceticus* recombinant host cells lacking one or more of these genes that are useful in the production of useful compounds.

The cells are useful in production in a variety of ways. First, certain cells make a useful FK-520-related compound merely as a result of inactivation of one or more of the 10 FK-520 biosynthesis genes. Thus, by inactivating the C-31 O-methyltransferase gene in *Streptomyces hygroscopicus* var. *ascomyceticus*, one creates a host cell that makes a desmethyl (at C-31) derivative of FK-520. Second, other cells of the invention are unable to make FK-520 or FK-520 related compounds due to an inactivation of one or more of the 15 PKS genes. These cells are useful in the production of other polyketides produced by PKS enzymes that are encoded on recombinant expression vectors and introduced into the host cell.

Moreover, if only one PKS gene is inactivated, the ability to produce FK-520 or an 20 FK-520 derivative compound is restored by introduction of a recombinant expression vector that contains the functional gene in a modified or unmodified form. The introduced gene produces a gene product that, together with the other endogenous and functional gene 25 products, produces the desired compound. This methodology enables one to produce FK-520 derivative compounds without requiring that all of the genes for the PKS enzyme be present on one or more expression vectors. Additional applications and benefits of such cells and methodology will be readily apparent to those of skill in the art after consideration of how the recombinant genes were isolated and employed in the construction of the 30 compounds of the invention.

The FK-520 biosynthetic genes were isolated by the following procedure. Genomic DNA was isolated from *Streptomyces hygroscopicus* var. *ascomyceticus* (ATCC 14891) 35 using the lysozyme/proteinase K protocol described in Genetic Manipulation of *Streptomyces* - A Laboratory Manual (Hopwood *et al.*, 1986). The average size of the DNA was estimated to be between 80 - 120 kb by electrophoresis on 0.3% agarose gels. A library was constructed in the SuperCosTM vector according to the manufacturer's instructions and 40 with the reagents provided in the commercially available kit (Stratagene). Briefly, 100 µg of

genomic DNA was partially digested with 4 units of *Sau3A* I for 20 min. in a reaction volume of 1 mL. and the fragments were dephosphorylated and ligated to SuperCos vector arms. The ligated DNA was packaged and used to infect log-stage XL1-BlueMR cells. A library of about 10,000 independent cosmid clones was obtained.

5 Based on recently published sequence from the FK-506 cluster (Motamedi and Shafiee, 1998, *Eur. J. Biochem.* 256: 528), a probe for the *fkbO* gene was isolated from ATCC 14891 using PCR with degenerate primers. With this probe, a cosmid designated pKOS034-124 was isolated from the library. With probes made from the ends of cosmid pKOS034-124, an additional cosmid designated pKOS034-120 was isolated. These cosmids 10 (pKOS034-124 and pKOS034-120) were shown to contain DNA inserts that overlap with one another. Initial sequence data from these two cosmids generated sequences similar to sequences from the FK-506 and rapamycin clusters, indicating that the inserts were from the FK-520 PKS gene cluster. Two *EcoRI* fragments were subcloned from cosmids pKOS034-124 and pKOS034-120. These subclones were used to prepare shotgun libraries by partial 15 digestion with *Sau3AI*, gel purification of fragments between 1.5 kb and 3 kb in size, and ligation into the pLitmus28 vector (New England Biolabs). These libraries were sequenced using dye terminators on a Beckmann CEQ2000 capillary electrophoresis sequencer, according to the manufacturer's protocols.

20 To obtain cosmids containing sequence on the left and right sides of the sequenced region described above, a new cosmid library of ATCC 14891 DNA was prepared essentially as described above. This new library was screened with a new *fkbM* probe isolated using DNA from ATCC 14891. A probe representing the *fkbP* gene at the end of cosmid pKOS034-124 was also used. Several additional cosmids to the right of the 25 previously sequenced region were identified. Cosmids pKOS065-C31 and pKOS065-C3 were identified and then mapped with restriction enzymes. Initial sequences from these cosmids were consistent with the expected organization of the cluster in this region. More extensive sequencing showed that both cosmids contained in addition to the desired sequences, other sequences not contiguous to the desired sequences on the host cell chromosomal DNA. Probing of additional cosmid libraries identified two additional 30 cosmids, pKOS065-M27 and pKOS065-M21, that contained the desired sequences in a contiguous segment of chromosomal DNA. Cosmids pKOS034-124, pKOS034-120, pKOS065-M27, and pKOS065-M21 have been deposited with the American Type Culture Collection, Manassas, VA, USA. The complete nucleotide sequence of the coding

sequences of the genes that encode the proteins of the FK-520 PKS are shown below but can also be determined from the cosmids of the invention deposited with the ATCC using standard methodology.

Referring to Figures 1 and 3, the FK-520 PKS gene cluster is composed of four open reading frames designated *fkbB*, *fkbC*, *fkbA*, and *fkbP*. The *fkbB* open reading frame encodes the loading module and the first four extender modules of the PKS. The *fkbC* open reading frame encodes extender modules five and six of the PKS. The *fkbA* open reading frame encodes extender modules seven, eight, nine, and ten of the PKS. The *fkbP* open reading frame encodes the NRPS of the PKS. Each of these genes can be isolated from the cosmids of the invention described above. The DNA sequences of these genes are provided below preceded by the following table identifying the start and stop codons of the open reading frames of each gene and the modules and domains contained therein.

	<u>Nucleotides</u>	<u>Gene or Domain</u>
15	complement (412 - 1836)	<i>fkbW</i>
	complement (2020 - 3579)	<i>fkbV</i>
	complement (3969 - 4496)	<i>fkbR2</i>
	complement (4595 - 5488)	<i>fkbR1</i>
	5601 - 6818	<i>fkbE</i>
20	6808 - 8052	<i>fkbF</i>
	8156 - 8824	<i>fkbG</i>
	complement (9122 - 9883)	<i>fkbH</i>
	complement (9894 - 10994)	<i>fkbI</i>
	complement (10987 - 11247)	<i>fkbJ</i>
25	complement (11244 - 12092)	<i>fkbK</i>
	complement (12113 - 13150)	<i>fkbL</i>
	complement (13212 - 23988)	<i>fkbC</i>
	complement (23992 - 46573)	<i>fkbB</i>
	46754 - 47788	<i>fkbO</i>
30	47785 - 52272	<i>fkbP</i>
	52275 - 71465	<i>fkbA</i>
	71462 - 72628	<i>fkbD</i>
	72625 - 73407	<i>fkbM</i>
	complement (73460 - 76202)	<i>fkbN</i>
35	complement (76336 - 77080)	<i>fkbQ</i>
	complement (77076 - 77535)	<i>fkbS</i>
	complement (44974 - 46573)	CoA ligase of loading domain
	complement (43777 - 44629)	ER of loading domain
	complement (43144 - 43660)	ACP of loading domain
40	complement (41842 - 43093)	KS of extender module 1 (KS1)
	complement (40609 - 41842)	AT1
	complement (39442 - 40609)	DH1
	complement (38677 - 39307)	KR1
	complement (38371 - 38581)	ACP1

	complement (37145 - 38296)	KS2
	complement (35749 - 37144)	AT2
	complement (34606 - 35749)	DH2 (inactive)
	complement (33823 - 34480)	KR2
5	complement (33505 - 33715)	ACP2
	complement (32185 - 33439)	KS3
	complement (31018 - 32185)	AT3
	complement (29869 - 31018)	DH3 (inactive)
	complement (29092 - 29740)	KR3
10	complement (28750 - 28960)	ACP3
	complement (27430 - 28684)	KS4
	complement (26146 - 27430)	AT4
	complement (24997 - 26146)	DH4 (inactive)
	complement (24163 - 24373)	ACP4
15	complement (22653 - 23892)	KS5
	complement (21420 - 22653)	AT5
	complement (20241 - 21420)	DH5
	complement (19464 - 20097)	KR5
	complement (19116 - 19326)	ACP5
20	complement (17820 - 19053)	KS6
	complement (16587 - 17820)	AT6
	complement (15438 - 16587)	DH6
	complement (14517 - 15294)	ER6
	complement (13761 - 14394)	KR6
25	complement (13452 - 13662)	ACP6
	52362 - 53576	KS7
	53577 - 54716	AT7
	54717 - 55871	DH7
	56019 - 56819	ER7
30	56943 - 57575	KR7
	57710 - 57920	ACP7
	57990 - 59243	KS8
	59244 - 60398	AT8
	60399 - 61412	DH8 (inactive)
35	61548 - 62180	KR8
	62328 - 62537	ACP8
	62598 - 63854	KS9
	63855 - 65084	AT9
	65085 - 66254	DH9
40	66399 - 67175	ER9
	67299 - 67931	KR9
	68094 - 68303	ACP9
	68397 - 69653	KS10
	69654 - 70985	AT10
45	71064 - 71273	ACP10

1 GATCTCAGGC ATGAAGTCCT CCAGGCCAGG CGCCGAGGTC GTGAACACCT CGCCCTGTGCT
 5 TGTACGGACC ATTTCACTCA CGGGCGATTG CGGAACCAAG TCATCCGGAA TAAAGGGCGG
 10 TTACAAGATC CTCACATTGC GCGACCGCCA GCATACGCTG AGTTGCCTCA GAGGCAAACC
 15 GAAAGGGCGC GGGCGGTCCG CACCAGGGCG GAGTACGCGA CGAGAGTGGC GCACCCGGCG
 20

241 ACCGTCACCT CTCTCCCCCG CCGGCGGGAT GCGCGCGGTG ACACGGTTGG GCTCTCTCG
 301 ACCTGAACA CCCGCGCGGT GTGGCGTCGG GGACACCGCC TGGCATCGGC CGGGTGAACGC
 361 TAGGGGGAGG GCGTACGGCG GCGGTGGCTC GTGCTACCGG CGGCGGGCG STCATCCGTC
 421 GAGACGGCAC TCAGCGAGCA GGGACGCGCTG STCGCGACCT GCGGGCCGGA CGACCGCTG
 481 GTTCCGGCG GGGCGGTGG CGGTGGTGA CGACGGTGG AGGGCGGTGA AGGCTGAGCG
 541 GTGACACGGC AGCAAAGGCC GGAGTCGGTC GGGGAAGGTG TCCACGAGGG CGTCGGTGTG
 601 CGTGCCTCGC TCGATCGGGT AGTAGCGGTG CGCGCGCGCA GGCAGCTGCC GGACATACCG
 661 GCCTACACGT CGGAGCCCG CGGGCAGGCA CGAGGACGTC GAGAGTGCCT CGATGGTGT
 721 CACCGGCTTG CGGATAACGAC CGGTCAACGC GATGCGTGG ACGGCCGGT CGACGCCG
 781 GGAGCGGGTG GCGTAGTCGT AGTCGGCATC CGACGCCGGG ACCGTCGGG GGGCGCAATA
 841 CGGTGTGCCG GCTTCCTTCT CCCCATCGAA GCCGGGGTCG AACTCCTCGC GGTAGACGCG
 901 CTCCGTCAGA TCCCAGTAGA CCTCGTGGTG GTACGGCCAC AAGAACCTCGG AGTCGGCCG
 961 GA-CCCGGGCG CGGAGCAGCG CCTCGCGCGC CTGGCCGGCT GCGGGGCCG CTGCCGCTA
 1021 GGTGGGGTAG TCGCGCAGGG CGGCCGGCAG GAAGGGTGAAG AGGTTGGGAC CCTCCGCG
 1081 CCACAGGGTG CCTTCCCAGT CGACTCCTCC GTCGTACAGC TCGGGATGGT TCTCCAGCTG
 1141 CCAGCGCACG AGGTAGCCGC CGTTGGACAT CCCGGTGACC AGGGTGCGCT CGAGCGGCCG
 1201 GTGGTAGCGC TGGGCGACCG ACGGCGGGC GGCCCCGGTC AGCTGGGTGA GGCGGGTGT
 1261 CCACTCGGCC ACGGCGTCGC CGGGCCGGGA GCCATCACGG TAGAACGCGG GGCGGGTGT
 1321 GCCCTTGTGCG GTGGCGCGT AGGCGTAACC CGGGCGAGC ACCCAGTCGG CGATGGCCCG
 1381 GTCGTTGGCG TACTGCTCGC GTTACCGGG GGTGCCGGCC ACGACCAGGC CACCGTTCCA
 1441 GCGGTCGGGC AGCCGGATGA CGAACTGGGC GTCGTGGTTC CACCCGTGGT TGGTGTGGT
 1501 GGTGGAGGTG TCGGGGAAGT AGCCGTCGAT CTGGATCCCG GGCACCTCCGG TGGGAGTGGC
 1561 CAGGTTCTTG GCGTCAGCC CTGCCCAGTC CGCCGGSTCG GTGTGGCCGG TGGCCGCCGT
 1621 TCCCGCCGTG GTCAGCTCGT CCAGGCAGTC GGCGTGTGA CGTGCCGCCG CGGGGACACG
 1681 CAGCTGGGAC AGACGGGCAG AGTGACCGTC CGGGGCATCG GGAGCAGGCC GGCGCTGGC
 1741 CGGTGAGGGG AGCAGGACGG CGACTCGGGC CAGGGTGAAG GCGCCGAGGC CGGTGCGTCT
 1801 TCTCGGGGCC CGTCCGACAC CGAGGGGCAG AACCATGGAG AGCCTCCAGA CGTGCAGATG
 1861 GATGACGGAC TGGAGGCTAG GTCGCGCAGC GTGGAGACGA ACATGGGTGC GCCCGCCATG
 1921 ACTGAGGCC CTCAGAGGTG GGCGCGCGCC ATGACGGCG CGGGACCGCG GGCCTCCGG
 1981 GCGGTGCCG CGGGCGCCA CGGGTCCGG GTCCCCGGT CAGGGACAGG TGTGTTCGC
 2041 GACGGTGAAG TAGCCGGTCG GCGACTCTTT CAAGGTGGTC GTGACGAAGG TGTGTTACAG
 2101 GCCCATGTC TGGCCGGAGC CCTTGGCGTA GGTGTAACCG GCGCTCGTC TGGCGCGGCC
 2161 CGCTGGACG TGAGCGTAGT TGGCGCGGT CCAGCAGACG GCGTGGCAC CGGTCGTC
 2221 CGCGGTGACC CGGCCCCAGA CGGGTCCCGC CTTGCGTGC GCGTCCCAGC CGGGGACCGC
 2281 GTAGGTGTC GATGTGCCG CCTCAGGGC GGTGTCGGT TACGACGTCG TGGCGGACGT
 2341 GGTGATCTGG GCACCGTCG GGTGGACGGC GTAGTCGGT GCGCCGTCGA CGGGTTCC
 2401 GGTCAAGGCTG ATGGTGGTGT CGGTGGCGCC GGTGGCGGCC AGGGCGGACG GAGCGGGCAG
 2461 CGAACCGGGG TCGGAGGGG ATCCGCTCAG CGCGAAGAAC TCGGTGATCC AGTAGCTGGA
 2521 ACAGATCGAG TCCAGGAAGT AGGCGGGCAGC GGTGCTCCGG CACTGCTGTG CTCCGGTGC
 2581 GGGATCGACC GGGGTGCCGT GCCCAGTCG CGGCACCCGG TTCACCTCCA CGGACCG
 2641 TCCCTCCGCG GCCAGGTACT CCTCGTCCGG GTGGAGGTTG GGGCCGATCA CGGAGGTACG
 2701 GTCCGGCGTC TGGGACACGC CGTGCACAGC GGTCCACTGG TCGCGCAACT CGTCGGCGT
 2761 GCGCGCGCG ACGGTGGTGT CCTTGTGCGC GTGCCAGATG GCCACGCGC GGCACGGG
 2821 CGACCAACGAG GGGTAGCCGT CACGGACCCG CGCGCCCCAC TGGTCCCGGG TCAGGTCGG
 2881 CCCGGGTTTC ATGCACAGGT ACGCGCTGCT GACGTCGGT GCACAGCGCA AGGGCAGGCC
 2941 GGCACGACCG GCGCCGGCCT GGAAGACGTC CGGATAAGGTG GCGAGCATCA CGACGTC
 3001 GGCACCGCCG GCGGACAGCC CGGTGATGTA GGTGCGCTGG GGGTCCCGCC CGTAGGGCGA
 3061 GACGGTGTGA CGGGCCATCT GCGGGATCGA CGCGGCTTCG CCCTGGCCCC TGGGTTCTC
 3121 GCTGCTCTGG AACCAAGTGA ACCACCTGTT CGCGTTGTTG GACGACGTTG TCTCGCGAA
 3181 CACGAGCAGG AAGCCATAGC GGTCCCGGAA TGAGAGCAGG CGGGAGTTGT CGGCGTAGCC
 3241 CTGGCGTCG TGGGTGCAAC CGTGCAGGGC GAACACCAAC GCGGGCTCCG CGGGCAGGG
 3301 CGCGGGCCGG TAGACGTACA TGTTCAGCCG GCGGGGTTC GTGCCGAAGT CGCGCACCTC
 3361 GGTCAAGGTCC GCCTTGGTCA GACCGGGCTT GGCAGGGCCC GCGCGGGCGT GGGCCGCG
 3421 CGCGGGGCCG AGCAGGGCG CTCCGAGTAC GAGGGCCACG ACGGCCACGA GACGGGTGAG
 3481 CACCCCCCGC CGTCCCGGAC GCGACAACGA CCCGACCGGC GCGGAGGAGG AGAGGGGAA
 3541 CAGCGGGGTG AGGATCCCC GGAACGGCGG CGGCTCGATG GCGGCTCCCT CGATGTCGTG
 3601 GGGGGACAC GGAGGGCTCC CTGACGTCA TCAAGTGGAG CGCCCCGGTG CGGGCACCG
 3661 TAGGGGTGCT TCAACCGCA ACGGTATGGC CGGAGCACC ACACCCCGCA CGCGCGATG
 3721 TGCGCCCGGA CGGATTGTGT CGCCTTGCAG AATCTGATAC CGGGACGCGA CGAACGCC
 3781 ACCCGACACG GGTAGGGCGT CATGGTGTCC GACTCGGCCG GTGGCGCTTG CCTGCCCTGG

3841 ACGGACCGGG CGTCGGCGGA CCGGGCGTCG GCGGGCTGGG CCGTATGGCG GCCGAGGACG
 3901 CCAGCCCGT GGGGGCGGCC CGCCCAAGTG CAGTACGCG ACCGTGGCCG GCAGGAGGGC
 3961 CGGACCGGTG AGTGCAGTCC CGCGGCCCTG CGGGACCGCT CGTCCCAGAC GGGTCCACC
 4021 GCGGCGAACG CGGCTCCGTG TCCGCGGGCG TAGACCATCA GTGTCGGCTC GAGGTGATE
 4081 ACGATGACAC CGTCCTGGTT GTAGCCGATG GTGCGCACGC TGATGATGCC TAGTCAGGT
 4141 CGGCTGGCGG ACTCCGGGT GTTCAGGACC TCGGACTGCG AGTAGATGGT GTCCGCGCTCG
 4201 AAGACCGGGT TCGGCAGCT GACCCGGTCC CAGCCGAGGT TCGCCATCAC ATGCTGGGAG
 4261 ATGTCGGTGA CGCTCTGCCG GGTGACCAAGG GCGAGGGTGA AGGTGGAGTC CACCAAGCGG
 4321 TTGCCCCAGG TGGTGGCCGC CGAGTAGTGG CGGTGCGAAGT GCGAGCGGC GGTGTTCTCG
 4381 GTCAGGAGC TGAGCCAGGA GTTGTGGTC TCCAGGACCG TGCGGCCAG GGGTGGCGG
 4441 TACACGTGCG CGGTGGTGA GTTCCTCGAAG TAGCGGCCCT GCCAGCCCTC GACCACAGCG
 4501 GTGCGGGTGG CGTCCTGGTC CGGGTTCTCA GTCGTCATGG CGCTCATCTC GGGAAAGTCCC
 4561 CGGTCCGCTG TGAAATGCCG AACCTTCACC GGGCTCATAAC GTGCGGCCGA TGAGCCCTGG
 4621 ACCGTACGTA GTCGTAGAAC CTCGCCACCA CTGGCGCGC TGTTCTCCG GCGAGTGTGA
 4681 CCACGCCGAC CGTGCGCCGC GCCTGCGGGT CGTCCAGCAG CACGGCGACG GCGTGGTCAC
 4741 CGGGCCCGGA CGGGCTGCCG GTGAGGGGG CGACGGCCAC ACCGAGGGCG GCGGCCGACCA
 4801 GGGCCCGCAG CGTGCTCAGC TCGGTGCTCT CCAGGACGAC CGCGGGCACG AATCCGGCG
 4861 CGGGCGACAG CCGGTGGTG ATCTGGCGCA GTCCGAAGAC CGGCTCCAGT GCCACGAACG
 4921 CCTCATCGGC CAGCTCCGCG TCCGCAACCC GGCGGCGTCT GGCCAGCCGG TGTCGGGTG
 4981 GGACGAGCAG GCACAGTGCC TCGTCCCGCA GTGGTGTCCA CTCCACATCG TCCCCGGCG
 5041 GTCGTGGGCT GGTCAAGCCCC AGGTCCAGCC TGCTGTTGCG GACGTGTCG ACCACGGCGT
 5101 CGGGCGCTG CGCGCGCAGT TCGAAGGTGG TGCGGGAGC GAGCGGGCGG TACCCGGCGA
 5161 GGAGGTGGG CACCAAGCCAG GTGCCGTAGG AGTCAAGGAA ACCCAGTGCC ACGGTGCGCG
 5221 GTCGGGGTC GATCAGGGCG GTGATGCGCT GTCGGCGCC GGAGACCTCA CTGATCGCG
 5281 GCAGGGCGTGG CGCGCGGAAG ACCTCGCCG ACTTGTGAG CGGGAGCCGG TTCTGGTGC
 5341 GGTCGAACAG CGGCACGCC ACTCGTCGCT CCAGCGCCG GATGGCCCTG GACAGGGTCG
 5401 GCTGGGAGAT GTTGAGCGT TCCGCGGTGA TCGTCACGTG CTCGTGCTCG GCCAAGGCCG
 5461 TGAACCACTG CAACTCCGT ATCTCCATGCA AGGGACTATA CGTACCGGGC ATGGTCCCTGG
 5521 CGAGGTTTCG TCATTCACA GCGGCCGGC GGCGGCCAC AGTGAAGTCT CACCAACCCAG
 5581 GACCCCATGG GAGGGACCCC ATGTCAGCAGC CGCATCCTCG CCTGAAACAG GAAACGCCCG
 5641 CGGGCCCGT GTCGGTCTG CTCGTGGTTT CTTTGGAGCA GGGCGTCGCC GCTCCGTTCG
 5701 CCACCCGCCA CCTGGCGGAC CTGGCGGCC CGTGTCACTCAA GATCGAACGC CCCGGCAGCG
 5761 GCGACCTCGC CGCGGCTAC GACCGCACGG TGCGTGGCAT GTCCAGCCAC TTGGTCTGGC
 5821 TGAACCGGGG GAGGGAGAG GTCCAGCTCG ATGTGCGCTC GCGGGAGGGC AACCGGCACC
 5881 TGCAAGCCTT GGTGGACCGG GCCGATGTCC TGGTGAGAA TCTGGCACCC GGCGCCGCGG
 5941 GCCGCCTGGC ATCGGCCACC AGGTCTCGC GCGGAGCCAC CGAGGCTGAT CACCTGCGGA
 6001 CATATCCGGC TACGGCAGTA CCGGCTGCTA CGCGGACCG CAAGGGCTAC GACCTCTGG
 6061 TCCAGTGCAGA AGCGGGCTG GTCTCCATCA CGCGCACCCC CGAGACCCC TCCAAGGTGG
 6121 GCCTCTCCAT CGCGGACATC TGTGCGGGGA TGTACCGCTA CTCCGGCATC CTCACGGCCC
 6181 TGCTGAAGCG GGGCCGCCACC GGGCGGGCT CGCAGTTGGA GGTCTCGATG CTCGAAGGCC
 6241 TCGGTGAATG GATGGGATAC GCCGAGTACT ACACCGCTA CGGCGGCACC GCTCCGGCCC
 6301 GCGCCGGCGC CAGCCACCGC AGGATCGCCC CCTACGGCCC GTTCACCAAG CGCGACGGGC
 6361 AGACGATCAA TCTCGGGCTC CAGAACGAGC GGGAGTGGGC TTCTTCTGC GGTGCGTGC
 6421 TACAACGCCCG CGGTCTCTGC GACGACCCCG GCTTTCCGG CAACGCCGAC CGGGTGGCGC
 6481 ACCGCCACCGA GCTGACGCC CTGGTGAGCC AGGTGACGGG CACGCTCACCG GGCAGAGAAC
 6541 TGGTGGCGCG GCTGGAGGAG CGTGCAGTCG CCTACCGCACG CGAGCGCAC GTCGGGAGT
 6601 TCAGCGAACCA CCCCCAACTG CGTGACCGTG GACGCTGGGC TCGTTCGAC AGCCCGGTCG
 6661 GTGCGCTGGC GGGCCTGATC CCCCCGGTCA CCTTCCACGG CGAGCACCCC CGGGCGCTGG
 6721 GCCGGGTCCC GGAGCTGGC GAGCATACCG AGTCCGTCCT GGCCTGGCTG GCGCGCCCC
 6781 ACAGCGCCGA CGCGAAGAG GCGGCCATG CGAATGAAC TCAACGGAGT CCTGATCCTG
 6841 GCCGCCGTGT CCTGCTCGC CGCGTACGG GGGCTGAACA TGGGCTGTG CCTCGCTGGTC
 6901 GCCACCTTTC TGCTCGGGGT GGTCGCACTC GACCGAACGC CGGACGAGGT GCTGGCGGGT
 6961 TTCCCCGCGA GCATGTTCTT GGTGCTGGTC GCGTCACGT CCTCTTCGG GATCGCCCGC
 7021 GTCAACGGCA CGGTGGACTG GCTGGTACGT GTCGCGTGC GGGCGGTGGG GGCCCGGGTG
 7081 GGAGCCGTCC CCTGGGTGCT CTCGCGCTG CGGGCACTGC TCTGCGCGAC AGGCACGGCC
 7141 TCGCCCGCGG CGGTGGCGAT CGTGGCGCCG ATCAGCGTCG CGTTCGCGCGT CAGGCACCGC
 7201 ATCGATCCGC TGTACGCCGG ACTGATGGCG GTGAACGGGG CCGCAGCCCG CAGTTCCGC
 7261 CCCTCCGGGA TCTCGGGCGG CATCGTCCAC TCGGCGCTGG AGAGAACCA TCTGCCCGTC
 7321 AGCGCGGGC TGCTCTCGC AGGCACCTTC GCCTCAACCG TGGCGGTGCG CGCGGTGTC
 7381 TGGCTCGTCC TCGGGCGCAG GCGCCTCGAA CCACATGACC TGGACGAGGA CACCGATCCC

7441 ACGGAAGGGG ACCCGGCTTC CGGCCCCGGC GCGGAACACG TGATGACGCT GACCGCGATG
 7501 GCCCGCGCTGG TGCTGGGAAC CACGGTCCTC TCCCTGGACA CCGGCTTCTT GGCCTCACG
 7561 TTGGCGCGT TGCTGGCGCT GCTCTTCCC CGCACCTCCC AGCAGGCCAC CAAGGAGATG
 7621 GCCTGGCCCG TGGTGCCTGCT GGTATGC333 ATCGTGCACCT ACGTCGCCT GCTCCAGGA
 7681 CTGGCGATCG TGGACTCCCT GGGGAAGATG ATCGCGGCGA TCGGCACCCCG GCTGCTGGCC
 7741 GCCCTGGTGA TCTGCTACGT GGGCGGTGTC GTCTCGGCCT TCGCCTCGAC CACCGGGATC
 7801 CTGGTGGCCG TGATGCCGCT GTCCGAGCCG TTCTGAAGT CGGGTGCCAT CGGGACGACC
 7861 GGCATGGTGA TGGCCCTGGC GGGCGGGCG ACCGTGGTGG ACGGCAGTCC CTTCTCCACC
 7921 ATGGTGCTG TGGTGGTGGC CAACGCTCCC GAGCGCTGC GGGCGGGCGT GTACCAAGGGG
 10 7981 TTGCTGTGGT GGGGCGCCGG GGTGTGCGA CTGGCTCCC CGGCCGCCCTG GGCGGCCCTC
 8041 GTGGTGGCGT GAGCGCAGCG GAGCGGGAAAT CCCCTGGAGC CGGTTTCCCG TGCTGTGTCG
 8101 CTGACGTAGC GTCAAGTCCA CGTGCCGGGC GGGCAGTACG CCTAGCATGT CGGGCATGGC
 8161 TAATCAGATA ACCCTGTCCC ACACGCTGCT CGCTTACGTA CGGAAGGGTGT CCCTGCGGA
 15 8221 TGACGAGGTG CTGAGCCGGC TGCGCGCGA GACGGCCGAG CTGCGGGCG GTGGCGTACT
 8281 GCCGGTGCAG GCCGAGGAGG GACAGTTCTC CGAGTTCTG GTGCGGTTGA CGGGCGCGCG
 8341 TCAGGTGCTG GAGATCGGG ACGTACACCGG CTACAGCAGC CTCTGCCTGG CCCGGGGATT
 8401 GGCGCCCGGG GGGCGTGTGG TGACGTGCGA TGTCATGCCG AAGTGGCCCG AGGTGGGCGA
 8461 GCGGTACTGG GAGGAGGCCG GGGTTGCCGA CGGGATCGAC GTCCGGATCG GCGACGCCCG
 8521 GACCGTCCTC ACCGGGCTGC TCGACGAGGC GGGCGGGGG CGGGAGTCGT TCGACATGGT
 20 8581 GTTCATCGAC GCGGACAAGG CCGGCTACCC CGCCTACTAC GAGGCGGCCG TGCGCTGGT
 8641 ACGCCGCGGC GGGCTGATCG TCGTCGACAA CACGCTGTT TGCGGCCGG TGCGCACGA
 8701 AGCGGTGCAG GACCGGACA CGGTCGCGGT ACGCGAACTC AACGCGGCAC TGCGCACGA
 8761 CGACCGGGTG GACCTGGCGA TGCTGACGAC GGCGACGCGC GTCACCTGC TGCGAAACG
 8821 GTGACCGGGG CGATGTCGGC GGGCGTCAGC GTCAGCGTC TGCGCGCGG CCTCGCGGAG
 25 8881 GGCTCCAGAT GCAGGCCGTT GACGCCGGCG CGGGAAAGCGC CCGCACCTC GGACACGAG
 8941 GGGCAGTCGG AGTCCCGA GCGCGAAC CCGTAGGCGA TCTCCATCAT GCGGTTGCGG
 9001 TCCGTACGCC GGAAGTCCGC CACCAGGTGC GCCCCCGCGC GGGCGCCCTG GTCCGTGAGC
 9061 CAGTTCAAGA TCGTCGACCC GGACCGAAC GACACGACCC GGCAGGACGT GGCAGACAGT
 9121 TTCAGGTGCC ACGTCGACGG CTTCTTCTCC AGCAGGATGA TGCGACGCG GCGTGCAGGG
 30 9181 CCGAAGCGGT CGCCCATGGT GACGACGAGG ACCTCATGGG CGGGATCGGT GAGCACGCGC
 9241 GCAGGTGCGC GTCGGAGTAG TGCACGCCGG TCGCGTTCAT CTGGCTGGTC CGCACGCTCA
 9301 GTTCTCTGAC GCGGCTGAGT TCCCTCTCCC CGCGGGTGC GATCGTCATG GAGAGGTGCA
 9361 GCGAGCGAG GAAGTCTCG TCGGGACCGG AGTACGCCCT CGGGGCTGG TCGCGCGA
 9421 AACCCGCTG GTACATCAGG CGCGGCCGAC GCGAGTCGAC CGTGGACACC GGCAGGCTGA
 35 9481 ACTCCGGCAG CGACAGGAGC GTGGCCGCGT GTCGGCCGG GTAGCACCGC ACCTCGGGCA
 9541 GTTGAACGC CACCTGGCA CGCTCGCGG GCTGGTCGTC GATGAACGCG ATCGTGGTCG
 9601 GTGCGAAGTT CAGCTCCGTG CGATCTCGC GGACGGACTG CGACTCGGC CCCCATCCGA
 9661 TGCGGCCAG CACGAAGTAC TCCGCCAACAC CGAGGGCTTC CAGACGCTCC CACCGAGGT
 9721 CGTGGTCGTT CTTGTCGCC ACCGCCTGGA GGATGCCGCG GTCGTCGAGC GTGGTATCA
 40 9781 CCTCGCGGAT CTCGTCGGT AGGACCAACT CGTCGTCCTC CAGCACGGTG CCCGCCACAA
 9841 AGGTGTTGTC CAGGTCCCG ACCAGACACT TGACAATGGT CATGGCTGTC CTCTCAAGCC
 9901 GGGAGCGCCA GCGCGTGTG GGCCAGCATC ACCCGGCACA TCTCGCTGCT GCGCTCGATG
 9961 ATCTCCATGA GCTTGGCGTC GCGGTACGCC CGTTGACGA CGTGTCCCTC TCTCGCGCCT
 45 10021 GCCGACGCGA GCACCTGTGC GGCGGTGCGC GCCCCGGCG CGGCTCGTT GCGGGCGACG
 10081 TGCTGGCCA GGATCGTCGC GGGCACCATC TCGGGCGAGC CCTCGTCCCA GTGGTCGCTG
 10141 GCGTACTCGC ACACGCGGGC CGCGATCTGC TCCGCGGTCC ACAGGTGCGC GATGTGCCCC
 10201 GCGACGAGTT GGTGGTCGCC GAGCGGCCGG CGAACATGCT CCCGGGTCCG GGCAGTGGCC
 10261 ACCGGCGCGG TGCGGCAGGC CGCAGGATC CGACGCGAC CCCAGGCGAC CGACTTGC
 50 10321 CCGTAGGCCA GTGACGCCGC GACCAGCATC GGCAGTGACG CGCCGGAGCC GGCCAGGACC
 10381 GCGCCGGCG GCACACGCCA CTGGTCCAGG TGCAGATCGG CGTGGCCGG CGCGCGGAG
 10441 CGGGACGGCT CGGGACGCC CTCGACGCCGT ACGCCGGGG TGTGGCGGG CACGACCA
 10501 ACCGCACCGG AACCATCCTC CTGGAGACCG AAGACGACCA GGTGGTCCCG TAGGGCGCG
 10561 GCAGTCGTGC AGACCTTGTG GCGTCGACG ACAGCGGTGT CCCCGTCGAG CGAACCCCGC
 10621 GTCCGCATCG CGCACAGATC GCTGCCCGCC TGCGCTCAC TGAAGCGAC GGCAGCGAGT
 55 10681 TTCCCGCTGG TCAGTCCTT CAGGAAGGTC GCGCGCTGAC CGCGCTCGCC GAGCCGCTGC
 10741 ACGGTCCACG CGGCCATGCC CTGCGACGTC ATGACACTGC GCAGCGAACT GCAGAGGCTG
 10801 CCGACGTGTG CGGTGAACCTC GCGTTCTCC CGGCTGCCGA GTCCCGAGACC GCGCTCTC
 10861 CCCGCCACTT CGCGCAGAG CAGGCCGTGS CGCGCGAGCC GGACGAGCGAG GTCGCGCGCG
 10921 AGTCGCGCGG ACGTGTCCA CTGGCGGGCC CGGTACCGA CAAGGTGGT CAGCAGCGCG
 60 10981 TCACCGCTCAG GCATCGACGG CCCGCAGCCG GTGGACGAGT CGACCATGG ACTCGACGGT

11041 AC3GAAGTTC GCGAGCTGGA GGTCCGGGCC GGCGATCGTG ACGTCGAACG TCTTCTCCAG
 11101 GTACACGACC AGTTCCATCG CGAACAGCGA CGTGAGGCCG CCCTCCCGCA ACAGGTGCGG
 11161 GTCCACGGGC CAGTCCGACC TGGTCTTCGT CTTGAGGAAC GCGACCAACG CGTGCACGAC
 11221 GG3GTCTGCG TTGACGGGTG CGGTCTATGAG AACACCTTGT CGTATTGCTA 3AAGCCCCCG
 5 11281 CC3GTCTTGC GGCGTGGTG TCCCTCGCG ACCTTGCCTA GCAGCAGGTC ACAGGGGCCG
 11341 CT3CGCTCGT CGCGGTGCG TTTGTCTCGC ACCCACAGCG CGTCGACGAG GTTGTGATG
 11401 CC3ATCAGGT CGCGGTGCG CAGCGGCCCG GTCGGATGGC CGAGGCACCC CGTCATGAGC
 11461 GC3TCGACGT CCTCGACGGA CGCGGTGCGC TCCGTGACGA TCCCGCCCGC GTCGTTGATC
 11521 ATCGGGTGGGA GCAGCCGGCT CGTGACGAG CGGGCGCGT CCCGGACGAC ATCGGCTTGS
 10 11581 CGCCCCAGCG CGCGGAGCAG GTCCCCGGCG CGGGCCATGG CCTTCTCACC GTTCCGGGGT
 11641 CCCGGATCA CCTCGACCGT CGGGATCAGG TACGACGGGT TCATGAAGTG CGTCCGAGC
 11701 AGGTCTCGG GCCGGGCCAC GGAGTCGCG AGTTCTGCAA CGGGGATCGA CGACGTGTC
 11761 GT3ATGACCG GGATACCGGG CGCGGTGCG GAGACCGTGG CGAGTACCTC CGCCCTGACC
 11821 TCGCGTCTC CGACGACGCG CTCGATCACC GCGGTGGCCG TACCGATCGC GGGCAGCGCG
 15 11881 GACGTGGCCG TCCGACGAC ACCGGGGTCG GCCTCGGCCG GCCCGGCCAC GAGTTGTGCC
 11941 GTCCGAGTTC CGGTGGCGAT CGCGC3CCC CGCGCCGTAA GGATCTCTC GGACGTGTC
 12001 ACGAGTGTCA CGGGGACGCC GTGGCGCAGC GCGAGCGTGG TGATGCCGGT GCCCATCACT
 12061 CCCCGCCGA GCACGATCAG CTGGTGGTCC ACGCTGTTTC CTCCCTCCCG GTCAACCATG
 12121 GCAGCGAGTA CGGGTCGAGG ACGTCTTCCG GGGTCGACCC GATCGCGTCC TTGCGGCCGA
 20 12181 GGCCGAGTTC GTCGGCGAAG CGGAGCAGCA CGTCGAACGC GATGTGGTCC GCGAACGCGC
 12241 TCGCCGTCGA GTCGAGGACG CTAGGCTGT CCCGGTGGTC CGCCGCGGGT TCCGGTGGCC
 12301 CGCACAGGGC CGCCAGCGAC GGGCCGAGCT CGCGGTCCGG CAGTTGCTGG TACTCGCCCT
 12361 CGCGCGGGC CTGCCCCCGA TGGTCGACGC AGATGAACGC GTCGTCGAGC AGGGTCTTCG
 12421 GCAGTTCGGT CTTGCCCGC TCGTCGGCGC CGATGGCGTT CACATGCAGG TGCAGCAGCC
 25 12481 GCGGCTCGGC GGGCAGCACC GGGCCTTGC CCGAGGGCAC CGAGGTGACG GTGGACAGGA
 12541 CATCCGCGGC GGCGGCGGCC TCCGCGGGAT CGGTACCTT GACCGGCAGT CCGAGGAACG
 12601 CGATGCGGTC CGCGAACGAC CGCGGTGGC CGGGGTCGGT GTCGCTGACC AGGATCCGCT
 12661 CGATGGGAG GACCCCTGCTG AGCGCGTGC CGTGGTCAC CGCCTGTGCG CCCGGCCGA
 12721 TCAGCGTGA CGTGGCGCTG TCGGACCGGG CCAGCAGCCG GTCAGCGACG GCGCGACCG
 30 12781 CGCCGGTCCG CATCGGGT ATCACGCTG CGTCGGCGAG GGCGGTCAAGA CTGCGCTGT
 12841 CGTCGTCGAG CGCGGACATC GTGCCGACGA TCGTCGGCGAG CGGAAAGCGC GGATAGTTGT
 12901 GCGGACTGTA CGAAACCGTC TTGATGGTCA CGCCGACACC GGGGACCCCG TACGGCATGA
 12961 ACTCGATGAC CGGGGAATG TCGCCCCCGC GGACGAATCC GTGACCGGGC GGCCTCTCGG
 13021 CGAAACTCGCC GCGGGCGAGC GGGCGAACCG CGTCGTGCA CGCCTGTGAC AGCCGGTCCA
 35 13081 TCATCACGTC GCGGGCGATC ACGGAGAGAA TCCGCTTGTG GTCACGTTGG CGCAGGACCC
 13141 TGGTCTGCAT GTGTCACCTC CTTTCGTGG CGCGAGCTGT CTTGGTGGTG CCGCTCGGG
 13201 CGGCTTCCGT TCTCATCGCA GCTCCCTGTG GATGAGGTG AAAATCTCGT CCGCGTGC
 13261 GTCCGCGGAC AGCACGCCG CGGGCGTGT CGGGCGGGT TCCCGCCGCC AGCGGTTGAG
 13321 CAGGGCGTCC CGACGGGTT CGATCGCGT CGCGCAGTGC GCGCGCGAC GGGTAGTCGA AGACGAGCGT
 40 13381 AACGAGTGT TCCAGCCGGT CGAGCTGCG GAGCACCACG GTCACCGGGT CGTCCGGGGA
 13441 CAGCAGTTCA CCGATGCGGT CGCGCAGTGC GCGCGCGAC GGGTAGTCGA AGACGAGCGT
 13501 GGCAGGACAGT CGCAGACCGG TCCGCTCGTT GAGGCCGTT CGCAGCTGCA CGCGATGAG
 13561 CGAGTCCACA CGAGTTCCC GGAACGCCG GTCCTCCGGG ATGTCCTCCG GTTCCGCGTG
 13621 GCCCAGGACG GCGCTGCCCT TCTGCCGGAC GAGGGCGAGC AGGTGGTGGG GGCCTTCTG
 45 13681 CTCGTTGCGG GCGCTCCGGC GGGCGACGG CTTGGCGGG CCACGCGACA GCGGGAGGTC
 13741 CGGGCGCAGG TCGCCCGCCA CGCGGACGAC ACTGCCGTT CGGGTGTGGA CGGGCGCGTC
 13801 GTACATGCGC ATGCCCTGTT CGCGGTGAG CGCGCTCGCC CCACCCCTTCG GCATACGGCG
 13861 CCGGTGGCG TCGGTCAAGT CGCGCGTCAG GCCACTCGCC TGGTCCCACA GCCCCCACGC
 13921 GATCGACAGC CCTGGCAGCC CTTGTGCAAG CGGTGTTCG GCGAGCGCGT CGAGGAACGC
 50 13981 GTTCCCGGCC GCGTAGTGTG CCTGACCGGG GGTGCCAGC ACACCGGCCG CCGACGAGTA
 14041 GACGACGAAT CGGGCGAGGT CGGTGTCGCG GGTGAGCCGG TCGAGGTGCC AGGCGGCGTC
 14101 GGCCCTTGGGT TTGAGGACGG TGTCGATGCG GTCGGGGGTG AGGTTGTCGA GCAGGGCGTC
 14161 STCGAGGGTT CGGGCGGTGT GGAAGACGCG GGTGAGGGGT TGAGGGATGT GGGCGAGGGT
 14221 GGTGGCGAGT CGGTGGGGTG CGCCGACGTC GCAGGGGAGG TGGGTGCCGG GGGTGGTGTG
 55 14281 GGGGGGTGGG GTGCGGGAGA GGAGGTAGGT GTGGGGGTGG TTCAAGGTGGC GGGCGAGGAT
 14341 CCCGGCGAGG GTGCCGGAGC CGCCGCTGAT GACGACGCC CCCTCGGGGT CCAGCGGCCG
 14401 CGGGACCGTG AGGACGATCT TGCCGCTGCG CTGCGCGCG CTCATGGTCG CCAGCGCCTC
 14461 CGGGACCTGC CGCATGTCGT GCACCGTCAG CGCGAGCGGG TCGAGCACAC CGCGCGCGAA
 14521 CAGGGCGAGC AGCTCCGCGA TGATCTCCTT GAGCCGGTGT GGCCCCCGCGT CCATCAGGTC
 60 14581 GAACGGTCGC TGGACGGCGT GCCGGATGTC CGTCTTCCCC ATCTCGATGA ACCGGCCACC

14641 CGCGCGGAGC AGGCCGACGG ACGCGTCGAG GAGTCACCG GTGAGCGAGT TGAGCACCGAC
 14701 GTCGACCGGC GGGAACCGGT CGCGAACCG GGTGCTCGGG GAATCGGCCA GATGCGCTCC
 14761 GTCCAGGTCC ACCAGATGGC GCTTCGCGGC GCTGGTGGTC GCGTACACCT CGCGGCCAG
 14821 GTGCCCGCGG ATCTGCCGGG CGCGGAAACC GACACC3CCG STGGCCCGGT GGATCAGGAG
 5 14881 CTTCTCGCCG GGGCGCAGCC CGCGGAGGTC GACCAGGCCG TACCAACGCCG T33CGAAGG
 14941 GGTCAATCACG GACGCCGCT CGGGGAACGT CCAGCCGTCC GGCATCCGGC CGACATCGG
 15001 GTGGTCGGGC ATGACCGTGG GGCGAAGGCC GGTGCCGACG AGGCCGAAGA CGGGTGC
 15061 CGGTGCCAGA CGGGAGACGT CGCGGCCGGT CTCCAGGACG ATGCCCGCGG CCTCGCCGCC
 15121 GACGACGCCG TGACCGGGT AGGTGCCGAG CGCGATCAGC ACATCGCGGA AGTTGAGGCC
 10 15181 CGCCGCACGC ACACCGATCC GGACCTCGGC CGGGGCGAGG GGGCGCCGGG GCTCCGCCGA
 15241 GTGGCCCGG GTGAGGCCGT CGAGGGTGC CGTCCCGGCC GGCGGATCA GCCACGTGTC
 15301 GCTGTCCCGC ACGGTGAGCG GCTCCGGCAC CGGGGTGAGG CGGGCCGCCG CGAACCGGCC
 15361 CGCGCGCAGC CGCAGACGCCG GCTCGCCGAG TGCGACGCCG ATGCGCTGCT GCTCGGGG
 15421 GAGCGTGACG CGGGACTCGG TCTCGACGTG GACGAACCGG CGGGCTGCT CGGCCTGGG
 15481 GCGCGCAGC AGTCCGGCG CGCGCCGGT GGCAGGGCCC GCGGTGGTGT GCACGAGCAG
 15541 ATCCCCGCGG GAGCCGGTCA GGGCGGTCA CGAGCCGGTG GTGAGCGCAC GCGTCTCGG
 15601 CACCGGGTCA TCGCCATCAG CGGCAGGCAA CGTGATGACG TCCACGTGCG TCGGGGG
 15661 ATCCGTGGGT GCGGCACCT CGATCCAGGT GAGACGCATC AGGCCGGTGC CGACGGTGG
 15721 GGACAGCGGG CGGGTGCAGA CGTCCCGGAT CTCGGCGACG AGTTGGCCGG CGGAGTCGG
 20 15781 GACCGCGAGA CTCAGCTCGT CGCCGTCACG AGTGATCAGC GCTCGGAGCA TGGCCGAGCC
 15841 CGTGGCGACG AACCGGGCCC CCTTCCAGGC GAACGGCAGA CCCGAGGCC TGTGTCGG
 15901 CGTGGTGAGG GCGACGGCGT GCAGGGCCGC GTCGAGCAGC GCCGGATGCA CACCGAAACC
 15961 GTCCGCTCG GCGGCTGCT CGTCGGGCAG CGCCACCTCG GCATAACACGG TGTACCCATC
 16021 ACGCCAGGA GCGCGAACCC CCTGGAACGC CGACCGTAC TCATAACCGG CATCCCGAG
 16081 TTCGTATAG AACCCCCAGA CGTCGACGCCG CACGGCCGTG ACCGGCGGGC ACTGCGAGAA
 16141 CGGCTCCACA CGACACAC CGGGGGTGTG GGGGGTGTG GGGGTCAAGG TGCCGCTGGC
 16201 GTGCCGGGTC CAGCTGGCG TGGCTCGGT ACAGCGTGTG ACGGTACCCG GCCGGCGTCC
 16261 GGCTCATCA GCGCTTCCA CGGTACCCGA CACATCCACC GCTGCGGTCA CGGGCACCCAC
 16321 AAGGGGGGAT TCGATGACCA GCTCGTCCAC TATCCCGCAA CGGGTCTCGT CACCGGCCCC
 30 16381 GATGACCAGC TCCACAAACCG CGCTACCCGG CAGCAGGACC GTGCCCCGCA CGGGCTGATC
 16441 ACCCAGCCAG GGGTGAGTGC GCAATGAGAT CGGGCCAGTG AGAACAAACAC CACCATGTC
 16501 GCGGGCAGC GCTGTGACAG CGGCCAGCAT CGGATGCGCC GCACCCGTCA ACCCCGCCGC
 16561 CGACAGATCG GTGGCACCGG CGGCCTCCAG CGAGTACCGC CTGTGCTCGA ACGGTACGT
 16621 GGGCAGATCC AGCAGCCGTC CGGGCACCGG TTCGACCACC GTGTCCCAGT CCACTGCCGT
 35 16681 GCGCAGGGTC CACGCCCTGCG CCAACGCCGT CAGCCACCGC TCCCAGCCG CGTCACCGGT
 16741 CCGCAACGAC GCCACCGTGT GAGCCTGTC CATGCCCGC AGCAGCACCG GATGGGCACT
 16801 GCACTCCACG AACACCGACC CATCCAGCTC CGCCACCGCC GCGTCAACG CCACCGGACG
 16861 ACGCAGATTC CGGTACCACT ACCCCTCATC CACCGGCTCC GTCACCCAGG CGCTGTCCAC
 16921 GGTGACCCAC CACGCCACCG ACGCGGCCTT CCCTGCCACC CCCTCCAGTA CTTGGCCAG
 40 16981 TTCATCCTCG ATGGCTTCCA CGTGGGGCGT GTGGGAGGCG TAGTCGACCG CGATAACG
 17041 CACCCGCACG CCTTCGGCCT CATACCGCGC CACCACTCC TCCACCGCCG ACGGGTCCCC
 17101 CGCCACCAACC GTCGAAGCCG GGGCGTTACG CGCGCGATC CACACACCC CGACCAAGACC
 17161 GACCTCACCG GCGGGCAACG CCACCGAAGC CATCGCTCCC CGCCCGGCCA GTCGGCCGC
 17221 GATGACCTGA CTGCGCAATG CCACCCAGCG GCGGGCGTCC TCGAGGCTGA GGGCTCCGGC
 45 17281 CACGCCACGCC GCGCGATCT CGCCCTGGGA GTGTCCGATC ACCCGTCCCG GCACGACCCCC
 17341 ATGCCCTGCG CACAGCCGG CGAGGCTCAC CGCGACCGCC CAGCTGCCG GETGGACCAC
 17401 CTCCACCCGC TCCGCCACAT CGGGCGCGC CAACATCTCC CGCACATCCC AGCCCGTGTG
 17461 CGGCAGCAAC GCCTGAGCGC ACTCCCTCAT ACGCGCGCG AACACCGGG AGTGGGCCAT
 17521 GAGTCCACG CCCATGCCGA CCCACTGGGC GCCCTGGCG GGGAAAGACGA ACACCGTACG
 50 17581 CGGCTGGTCC ACCGCCACAC CGGTACCCCG GGCATC3CCC AGCAGCACCG CACGGTACCC
 17641 GAAGACAGCA CGCTCCCGCA CCAACCCCTG CGCGACCGCG GCCACATCCA CACCAACCCCC
 17701 GCGCAGATAC CCCTCCAGCC GCTCCACCTG CCCCCCGCAGA CTCACCTCGC CACGAGCCG
 17761 CACCCGCAAC GGCACCAACC CGTCAACAAAC CGACTCCCCA CGCGACGCC CAGGAACACC
 17821 CTCAAGGATC ACGTGCGCGT TCGTACCGCT CACCCCGAAC GACGACACAC CGCGATGCGG
 55 17881 TGCCCGATCC GACTCGGGGCC ACGGCCCTCGC CTGGTGGAGC AGCTCCACCG CACCGGCCGA
 17941 CCAGTCCACA TCGGACGACG GCTCGTCCAC ATGCAGCGTC TTGGCGGCCA TCCCGTACCG
 18001 CATGCCCATG ACCATCTGA TCACACCGGC GACACCC3CC GCGCCCTCGC CATGACCGAT
 18061 GTTCGACTTC AACGAACCCA GCAGCAGCGG AACCTCACGC TCGTACCGT ACGTGCCAG
 18121 AATGGCCTGC GCCTCGATGG GATCGCCCGAG CGTGTCCCC GTCCCGTGGC CCTCCACCCAC
 60 18181 GTCCACATCG CGGGCGCGCA GTCCGGCGTT CACCAACGCC TGCTGGATGA CACGCTGCTG

18241	GGACGGGCCG	TTGGGGCGG	ACAGCCGTT	GGAGGCACCG	TCCCTGGTCA	CCGGCGACCC	
18301	GCGGACGACC	GCGAGAACGG	TGTGTCCGTT	GCGCTCGGCG	TCGGAGAGCC	GCTCCAGCAC	
18361	AAGAACGCCG	GCGCCCTCCG	CCCAGCCGGT	GCCGTTGGCG	CGTCCCGCGA	ACGGCGGGCA	
18421	GCGGCCGTG	GGGGAGAGTC	CGCCCTGCTG	CTGGATTCC	ACGAACCCGG	TCGGGGTGCG	
5	18481	CATGACGGTG	ACACCGCCGA	CCAGCGCCAG	CGAGCACTCC	CCGTGGCGCA	GTGGTGTGCC
	18541	GGCCTGGTGC	AGCGCGACCA	GCGACGACGA	GCACGCCGTG	TCCACCGTGA	ACGGCGGTCC
	18601	CTGGAGCCCA	TAGAAGTACG	AGATCCGGCC	GGTGAGCACG	CTGGGCTGCA	TGCCGATCGA
	18661	GCGAACCCG	TCCAGGTCCG	CGCCGACGCC	GTACCCGTAC	GAGAAGGCAC	CCATGAACAC
	18721	GCCGGTGTG	CTGCCGCGCA	GTGTGCCCGG	CACGATGCC	GCGCTCTCGA	ACGGCTCCCA
10	18781	TGTCGTTTCC	AGCAGGATCC	GCTGCTGGGG	GTCCATGGCC	CGTGCCTCAC	GGGGGCTGTAT
	18841	GCGGAAGAAC	GCGGCATCGA	AGCCGGCGGC	GTGAGGAGGG	AAGCCGCCGC	GGTCCGTGTC
	18901	CGATCCGCCG	GTGAGGCCGG	ACGGGTCCC	GCCACGGTGC	GCCGGGAAGC	CGGTGACCGC
	18961	GTGCCCCC	CTGTCCACCA	TGCGCCACAG	GTCGTCGGGC	GAGGTGACGC	CCCCCGGCAG
15	19021	TCGGCAGGCC	ATGCCCACGA	TGGCCAGCGG	TTCGTCACGG	GTGCGGGCGG	CTGTGGGAAC
	19081	AGCGACCGGT	GCGGCACAC	CGACCAGAGC	CTCGTCCAAC	CGCGACGCCA	TGGCCCGCGG
	19141	CGTCGGGTAG	TCGAAGACAA	GCGTGGCGGG	CAGTCGGACA	CCGGTCCCGG	CGGGAGTCG
	19201	GTTCCGCAGT	TCGACGGCGG	TCAGCGAGTC	GATACCCAGT	TCCCTGAAGG	CCCGCTCCGC
	19261	GGACACGTCC	GCGCGTCCG	CGTGGCCGAG	CACCGCCGCC	GCCTTGTGCG	GGACCAGTGC
	19321	CAGCAGCGCG	GTGTCGGCT	CAGCGCCGGA	CATGGTGC	AGCCGGTCGG	CGAGCGGAAC
20	19381	GGCGGTGGCC	GCCGCCGGGC	GCGATAACGGC	GCAGCGCAGA	TGGCGAAAAA	CGGGCGATGT
	19441	GTGCGCGGTG	AGGTCCATCG	TGGCCGCCAC	GGCGAACGCG	GTGCGGGTTC	CGGGCGCGGC
	19501	TTCACAGCAGG	CGCATGCCA	CACCGGCCGA	CATGGGGCGG	AAACCGCCGC	GGGGGACACG
	19561	GGTGCCTGTTG	GTGCCGCTCA	TGCTGCCGGT	GAGTCCGCTG	TCATCGGCC	AGAGGCCCCA
	19621	GGCCAGCGAC	AGCGCGGGCA	GTCTTCGGC	ATGGCGCAGC	GTGCGAGTC	CGTCGAGGAA
25	19681	CCCGTTCGCC	GCGAGTAGT	TGCCCTGGCC	GCAGCGGCC	ATGATGCCG	CGACGGACGA
	19741	GTAGAGGACG	AACGAGCGCA	GGTCGCGTC	CGGGGTCAGC	TGTCGAGGT	GCCAGGCGCC
	19801	GTCGGCTTGC	GGGCGCAGTG	TGGTGGCGAG	CGCCTCCGGG	GTGAGTGC	TGGTCACGCC
	19861	GTCGTCGAGC	ACGGCTGCCG	TGTTGAAGAC	CGCCGTGAGC	GGCCTGCCG	CGGGCGCGAG
	19921	CGGGCGGGG	AGCTGGTCCC	GGTCGGCGAC	GTCACAGCGG	ATGTCGACAC	CGGGAGTGT
30	19981	CGCCGGCGGT	TCGCTCGCG	ACAGCAACAG	GAGGTGGCGG	GCGCCATGCT	CGGCGACGAG
	20041	ATGCCGGCGG	AGGAGACCTG	CCAGCACACC	CGAGCCGCCG	GTGATGACCA	CCGTGCCGTC
	20101	CGGGTCGAGC	AGCGGTTCGG	GGCTTCCGC	GGCGGGCGTG	CGGGTGAACC	CGGGCGCTTC
	20161	GTACCGGGCG	TCGGTACGC	GGACGTACGG	CTCGGGCAGT	TGCTGGCGG	CGGCCAGCGC
	20221	CTCGATGGGG	GTGTCGGTGC	CGGTCTCCAC	CAGCACGAAC	CGGCCCCGGG	GTCGGCCTG
35	20281	GGCGGACCGG	ACGAGGCCGG	CGACCGCTCC	TCCGACCGG	CCCGCGTCGA	TCCGGACGAC
	20341	GAGGGTGGTC	TCCGCAAGGC	CGTCCTCGGC	GATCACCCGG	TGCACTCGC	CGAGCACGAA
	20401	CTCGGTGAGC	CGGTACGTCT	CGTCGAGGAC	ATCCGCGCCC	GGTTCCGGGA	GCGGGAGAC
	20461	GATGTGGACC	GCGTCCGCAG	GACCGGGCCC	GGGAGTGGC	AGCTCGGTCC	AGGAGAGGCC
	20521	GTACAAGGAG	TTCCGTACGA	CGGCGCGTC	GCCGTCGACG	TTCACCGGTC	GCGGGTCAG
40	20581	CGCGCGACG	GTCACCAACG	GTTGGCCGAC	CGGGTCCGTC	GCATGCACGG	CAGGCCGTC
	20641	CGGGCCCTGA	GTGATCGTGA	CGCGCAGCGT	GGTGGCCCCG	GTGCGTGTGGA	ACCGCACGCC
	20701	GCTCCACGAG	AACGGCAGCC	GCACCTCCGC	TTCTGTTCC	GCGAGCAGCG	GCAGGCAGGT
	20761	GACGTGCAAG	GCCGCGTCGA	ACAGCGCCGG	GTGGACGCCA	TAGTGCGGCG	TGTGTCGCC
	20821	CTGTCGGCG	GCGATCTCCA	CCTCGCGTA	CAGGGTTTCG	CCGTCGCGCC	AGGCGGTGCG
45	20881	CAGTCGGCG	AACGCTGGC	CGTAGCTGTA	GCCGGCTCG	GCCAGCCGCT	CGTAGAACGC
	20941	GCTCACGTCG	ACCGCGTCG	CGCCCGGCCG	CGGCCACGCG	GGCGGCGGGG	CCGCGCGAC
	21001	GCTTCGGCG	CGGCCGAGGG	TGCCGCTGGC	GTGCGGGGTC	CAGCTGTCCG	TGCCCTCGGT
	21061	ACGGCGCTG	ACGGTCACTC	GCCGCCGTCC	GGCCTCATCG	GCCCCCTCGA	CGGTCAACGA
	21121	CACATCCACC	GCGCCGGTCA	CCGGCACAC	GAGCGGGGTC	TCGATGACCA	GTCATCCAC
50	21181	CACCCCGC	CCGGTCTCGT	CACCGGCCG	GATGACCA	TCCACAAACG	CCGTACCCGG
	21241	CAGCAGAAC	GTGCCCCCGA	CCGGCTGATC	AGCCAGCCAG	GGATGCGTAC	CGAACGAGAT
	21301	CGGGCCAGTG	AGAACAAACAC	CACCAACGTC	GTCGGCGGGC	AGTGTGTCGA	CGGGCGCCAG
	21361	CATCGGATGC	GGCGCCCCGG	TCAGCCCGGC	CGCGGACAGA	TGGGTGGCAC	CGGCCGCGCC
	21421	CAGCCAGTAC	CGCCTGTGCT	CGAACCGCGTA	GGTGGCGAGA	TGCGACAGCC	GTCCCGGCAC
55	21481	CGGTTGACG	ACCGTGTCCC	AGTCCACTGC	CGTGGCCAGG	GTCCACGCT	CGCGCAACGC
	21541	CGTCAGGCCAC	CGCTCCCAGC	CGCCGTCACC	GGTCCGCAAC	GACGCCACCG	TGTGAGCCTG
	21601	TTCCATCGG	GGCAGCAGCA	CCGGATGGGC	GCTGCACTCC	ACGAACACGG	ACCCCTCGAC
	21661	CTCCGCCACC	GCCGCCGTCCA	GCGCAGCGGG	GCGACGCAGG	TTCCGCTACC	AGTAGCCCTC
	21721	ATCCACCGGC	TCGGTCACCC	AGGCGCTGTC	CACCGTGGAC	CACCAAGGCCA	CCGACCCGGT
60	21781	CCCGCCGGAA	ATCCCCCTCCA	GTACCTCGGC	CAACTCGTCC	TCGATGGCTT	CCACGTGGGG

	21841	CGTGTGGGAG	CGTAGTCGA	CCSGCATAAG	GCGCACTCGC	ACGCCTTCGG	CCTCGTACCG
	21901	CGTCACCACT	TCTTCACCG	CGGACGGGTC	CCCCGCCACC	ACAGTCGAAG	ACGGGCCGTT
5	21961	ACCGGCCGCG	ATCCACACGC	CCTCGACCAG	GTCCACCTCA	CCGGCCGGCA	ACGCCACCGA
	22021	AGCCATCGCC	CCCCGCCCCG	CCAGCCGCC	GGCGATCACC	TGGCTGGCA	AGGCCACCGA
	22081	CGGGGCGGGCG	TCCTCAAGGC	TGAGGGCTCC	GGCCACACAC	CCGGCCGGCA	TCTCGCCCTG
	22141	GGAGTGTCCG	ACCACCGCGT	CCGGCACGAC	CCCATGCGCC	TGCCACAGCG	CGGGCAGGCT
10	22201	CACCGCGGAC	GCCCAGCTGG	CCGGCTGGAC	CACCTCCACC	CGCTCCGCCA	CATCCGGCC
	22261	CGCCAACATC	TCCCACAT	CCCAGCCCGT	GTGCGGCAAC	AACGCCCGCG	CACATCCTC
	22321	CATACGAGCC	GCGAACACCG	CAGAACACGC	CATCAACTCC	ACACCCATGC	CCACCCACTG
	22381	AGCACCCCTGC	CGGGGAAAGA	CGAACACCGT	ACGCGGCTGA	TCCACCGCCA	CACCCATCAC
15	22441	CGGGGCATCG	CCCAACAACA	CCGCACGGGT	ACCGAAGACA	GCACGCTCAC	GCACCAACCC
	22501	CTGCGCGACC	GCGGCCACAT	CCACACCACC	CCCGCGCAGA	TACCCCTCCA	GCCGCTCCAC
	22561	CTGCCCCCGC	AGACTCACCT	CACTCCGAGC	CGACACCGGC	AACGGCACCA	ACCCATCGAC
	22621	AGCCGACTCC	CCACCGGACG	GCCCCGGGAAAC	ACCCCTCAAGG	ATCACGTGCG	CGTTCGTACG
20	22681	GCTCACCCCCG	AAAGCGGAGA	CACCGGGCCCG	GGCGGGACGT	CCCGCGTCGG	GCCACGCCCC
	22741	CGCCTCGGTG	AGCAGTTCCA	CCGCGCCCTC	GGTCCAGTCC	ACATGCGACG	ACGGCTCGTC
	22801	CACATGCAGC	GTCTTCGGCG	CGATGCCATA	CCGCATCGCC	ATGACCATCT	TGATGACACC
	22861	GGCGACACCC	GCAGCCGCT	GCGCATGACC	GATGTTGAC	TTCAACGAAC	CCAGCAGCG
25	22921	CGGAACCTCA	CGCTCCTGCC	CGTACGTCCG	CAGAACCGCG	TGCGCCTCGA	TGGGATCGCC
	22981	CAGCGTCGTC	CCCCTCCCCTG	GCGCCTCCAC	CACGTCCACG	TCGGCGGGGG	CGAGCCCCCG
	23041	CTTGTGGAGG	GCCTGGCGGA	TGACGCGCTG	CTGGGAGGGG	CCGTTGGGTG	CGGAGATGCC
	23101	GTTGGAGGCG	CCGTCTGGT	TGACGGCGGA	GGAGCGGACG	ACCGCGAGGA	CGGTGTGTCC
	23161	GTTGCGCTCG	CGTGCAGGAGA	GCTTTTCGAC	GACGAGGACG	CCGGCCCCCT	CGGGGAAACC
30	23221	GGTGCCTGTC	GCGCGTCAG	CGAACGCCTT	GCACCGTCCG	TCCGGCGCGA	CGCCGCCCTG
	23281	CGGGGAGAAC	TCCACGAAGG	TCTGTGGTGA	TGCCATCACT	GTGACACCAC	CGACCAGCGC
	23341	CAGCGAGCAC	TCCCCGGTCC	GCAGCGCCTG	CCCGGCCCTGG	TGCAAGCGCGA	CCAGCAGCGA
	23401	CGAACACGCC	GTGTCGACCG	TGACCGCCGG	ACCCCTCCATG	CGGAAGAAAGT	ACGACAGCG
	23461	TCCGGCGAGC	ACCGCGGGCT	GTGTGCTGTA	GGCGCCGAAT	CCGCCCAGGT	CCGCGCCCGT
35	23521	GGCGTAGCCG	TAGTAGAAC	CGCCGACGAA	GACGGCGGTG	TCGCTGCCGC	GCAGGGGTGTC
	23581	CGGCACGATG	CGGGCGTGT	CGAGCGCCTC	CCAGCGATT	TCGAGGAGGA	TCCGCTGCTG
	23641	CGGGTCGAGT	GCGGTGGCCT	CGCGCGGACT	GATGCGGAAG	AACGCGGCAT	CGAAGTCGGC
	23701	GGCGCCCGCG	AGTGCGCCGG	CCCGCCCGGT	GGCGGACTCG	GGGGCGGGGT	CGAGCGCGGC
	23761	CACGTCCCAG	CCGCGGTGCG	TGGGGAAAGTC	CCGATCGCG	TCGGCGCCGT	CCGCGACGAG
	23821	CTGCCACAGC	TCTTCGGTG	AGGTGACGCC	CCCCGGCACT	CCGCAGGCCA	TGCCCACGAC
40	23881	GGCGAGCGGC	TGTTCCCG	CGCGCGCGAG	CGCGGTGTT	TCCCGGCCGA	GCTGCGCGTT
	23941	GTCCTTGACC	GACGTCCGCA	GCGCCTCGAT	CAGGTCGTT	TCCGGCATCG	CCTCATCCCT
	24001	TCAGCACGTG	CGCGATGAGC	CGCTCTGCGT	CCATGTCGTC	GAACAGTTG	TCGTCGGGCT
	24061	CCGCGTCGTC	GGTGCCTCGC	GGTGCCTGTT	CCGGTGGTT	ACCGCCGTC	GGGGTCCCGT
45	24121	TGTCGTCCGG	GGTCCCGTTG	ACGTCCGGGG	CCAGGAGGGT	CAGCAGATGA	CGGGGTGAGCG
	24181	CGCCGGCGGC	GGGATAGTCG	AAGACGAGCG	TGGCCGGCAG	CGGAATGCG	AGGGCCTCGG
	24241	AGAGCCGGTT	CGCGAGGCCG	AGCGCGGTGA	GCGAGTCGAC	CCCGAGGTTCC	TTGAACGCCG
	24301	TGGTGGCCGT	GACCGCCGCC	GGTGCCTGTT	GGCCCAGCAG	GGTGGCGGGG	GTGTCGCCGA
	24361	CGACGCCGAG	CAGCACCTGT	TCCCCTTCCT	TGTGGGGCAG	GTCCGGCAGG	CGTTCCAGCA
50	24421	GGGAGCCGCC	GTCGGTCGCG	GAGCGCCGGG	TGGGGCGCTG	GATCGGTCCG	CACAGCGGTG
	24481	ACGGGTCGCC	GGGCCCCGGG	GGGGCGGTG	CCACGACAC	GGCTTCCCCG	GTGGCGCACG
	24541	CGGCGTCGAG	GAGGTGGTC	AGCCGGTCCG	CCGCGGCCGT	GAACGCCACG	GCGGCAGGC
	24601	CTTGTGCCCG	GCGCAGGGTC	GCCAGGGCCT	GGAGCGGTCC	GGCCGCCCTCG	CCGGACGGAA
	24661	CGGCAGAAC	GAACGCCGTC	AGGTGAGGT	CGCGGGTCAG	GGGGTGCAGT	TCCCAGGCC
55	24721	ACTCGCCGT	GGCGTCCCG	TGGACGACC	CGGTCAACCG	GGTTTCCGGC	ACTGTGCCCG
	24781	GCTCGTACCG	GATCACTTCG	GGCCCGTGT	CGCCGAGGT	TCCGGCGAGT	TCCTCCGAAC
	24841	CGCCCGCAG	GAGGACGGTG	TCGCGTACG	AGGCCGCGGC	CGTGGTGGG	CGGGCGGGGA
	24901	CGAGGCGGGG	CGCTTCGAGG	CGCCCCGTG	CCAGGCGCAG	GTGCGGTTG	TCGAGGCGGG
	24961	AGAGGGCGGC	GGCGCGGCCG	GGGGTGAACCG	TGTCGTGCT	CTCCACGAGC	ACGAGCCGGC
	25021	CCGGTTCCGC	GGTGTGAGC	AGTGCAGGCGA	CGGCACCGGC	GACGGGCCCG	GCCTCGGCCG
60	25081	ACACCACCA	CGTGGCGCCG	GGGGTCCCTG	GGTCGTCCAG	TGCGGTACGG	ACCTCGTCGG
	25141	GACCGGATAC	CGGGACGACG	ATGACGTG	CGTGGCGTC	GTCGCCGAGG	TCGGTGTAC
	25201	GGGGGGCCGT	GGTGCCTGGG	GGCGCCGGG	CCCGGACGCC	GGTCCAGGTG	CGCCGGAACG
	25261	GGCGCACGTC	CCCCTCCGGG	CCCCTCGTGTG	GGGGGGCCG	GGTGTGAGC	GAGCCGATCT
	25321	GAGCCACCGG	CCGTCCCAGT	TCGTCGGCGA	GGTGCACGCG	GGCGCCGCC	TCGCCCTCGC
	25381	CGTGGACGAA	GGTACGCGC	AGTTTCGTTG	CGCCGCTGGT	GTGGACACGG	ACGCCGGTCA

35441	ACCGGAACGG	CAACCGTACC	CCCCCGTTCT	CGCGGGCCGC	GCCGATGCTG	CCCGCTTGCA	
26501	CGCGGGTGAC	GAGCAGCGCC	GGGTGCAGTG	TGTAGCGGGC	GGCGTCCCTG	CGAGGGCGC	
25561	CGTGGAGGGC	GAATTCGGCG	CAGACGGTGT	CTCCGTGGCT	CCACGCGGCG	GACATGCC	
25621	CGAAACTCGGG	CGCGAACTCG	TATCCCGCT	CGTCGAGTCG	CTGGTAGAAG	CGCGGACCT	
5	25681	CGACCGGTT	CGCGTGTCTG	GGCGGCCAGG	GCCCCGGCGT	GGTGGCCGGT	TGCGTGGTGG
25741	CGATGCCGGC	GAAGCCGGAG	GCCTGGCGGG	TCCATGTCCG	GTCGCCGTCC	GTCCGGCGT	
25801	GGACGCGCAC	GGCACGGCGT	CCGGTGTCTG	CGGGCGCGGC	GACGGTCACG	CGCACCTGGA	
25861	CGGCGCCGGT	GGCGGGCAGG	ACCAGCGGTG	TCTCGACGAC	CAGTCGTCTG	AGCAGGTCCC	
10	25921	AGCCTGCTC	GTCGGGCCG	CGTCCGGCCA	ATTCCAGGAA	GGCGGGTCCG	GGCACAGTA
25981	CGCGCCGTC	GACGGAGTGA	CCGGCCAGGC	ATGGGTGGGT	GGCCAGCGAG	AACCGGCCGG	
26041	TGAGCAGCAC	CTCGTCGGAG	TCGGGGAGCG	CCACCGACGC	GGCGAGCAGC	GGGTGGTCGA	
26101	CGCGTCGAG	TCCGAGGCCG	GAAGCGTCCG	TGCCGGCCGC	GGTCTCGATC	CAGTAGCGCT	
15	26161	CATGGTGGAA	GGCGTATGTG	GGCAGGTCTG	GTGCCGTGCG	CGTCGCGGGG	ACGACCGCCG
26221	CCCAAGTCGAC	GGGCACGCCG	GTGTGTGCG	CCTCGGCCAG	CGCGGTGAGC	AGCCGGTGGG	
26281	CTCCCCCGCC	GGGGCGGAGC	GTGGCGACGG	TGCGGCCGTC	GATCGCGGGC	AGCACGACGG	
26341	GGTGCACGCT	GACCTCGACG	AACACGGTGT	CACCCGGCTC	GGGGCAGCG	GTACACGGCCG	
20	26401	TGCGAAGCC	TACGGGTGG	CCCATGTTGC	GGAAACCAGTA	CTCGTCGTG	AGCGCCGCGT
26461	CGATCCAGCG	TTCGTCGGCG	GTGGAGAAC	ACGGGATCTC	GGGCGTGC	GAGGTGGTGT	
26521	CCCGACGAT	CCGCTGGAGT	TCGTCGTACA	CGGGGTGAC	GAACGGGGTG	TGGGTGGGCG	
26581	AGTCCACGGC	GATGCCGCGC	ACCCAGACGC	CGCGGGCCTC	GTAGTCGGCG	ATCAGCGTT	
26641	CGACGGCGTC	CGGGCGCCCG	GCGACGGTCG	TGGTGGTGGC	GCCGTTGCGG	CCCGCAGACC	
26701	AGACGCCGTC	GATCCGGGCG	GCATCCGCT	CGACGTCGGC	GGCCGGGAGC	GCGACCGAGC	
26761	CCATCGCGCC	CGTCCGGCG	AGTTCGCGCA	GGAGCAGGAG	AACGCTGCGC	AGCGCAGCGA	
26821	GGGGGGCACC	GTCCCTCCAGG	GTGAGCGCTC	CGCGACAC	GGCGCGGGCG	ATCTCGCCCT	
25	26881	GGGAGTGTCC	GATGACGGCG	TCCGGCGTA	CGCCCGCGGC	CTCCCACACG	GCGGCCAGCG
26941	ACACCATGAC	GGCCCAGCAG	ACGGGGTGC	CGACGTCGAC	CGGGCGGGTC	ACCTCCGGGT	
27001	CGTCGAGCAT	GGCGATGGGG	TCCCAGCCCC	TGTGCGGGAT	CAGCGCGTCG	GCGCATTTGGC	
27061	GCATCCCTGGC	GGCGAACACC	GGGGAGGCCG	CCATCAGTTC	GACGCCCATG	CCCGGCCACT	
30	27121	CGGGTCTT	TCCGGGAAAG	ACGAAGACGG	TGCGCGGCTC	GGTGAGCGCC	GTGCCGGTGA
27181	CGACGTCGTC	GTGAGCAGC	ACGGCGCGGT	CGGGGAAACG	CGTACGCTG	GCGAGCAGGC	
27241	CCCGGGCGAT	GGCGCGCGGG	TCTGGCGGG	GACGGCGGGC	GAGGTGCTCG	CGGAGTCGGC	
27301	GGACCTGGCC	GTGAGGGCC	GTGGCGGTCC	GGGCCGAGAC	GGGAGTGGT	GTGAGCGGGC	
27361	TGCGATCAG	CGGCTCACCG	GGCTTCGAGG	CCGACGGCTC	TCTGGCCCCG	GGCTCCCCGG	
35	27421	CCGGGTGGGC	TTCCAGCAGG	ACGTGGCGT	TGGTGGCGCT	GACCGGAAG	GAGGACACAC
27481	CGCGCGCCG	CGGGCGGTG	GTCTCGGGCC	AGGGCGGGC	ATCGGTGAGG	AGTCGACGG	
27541	CCCGGGCGT	CCAGTCGACG	TGCGAGGAGC	CGTGTCCAC	GTGAGGGTG	CGCGCAGGG	
27601	TGCCCTGCG	CATGGCGAGG	ACCATCTGA	TGACACCGC	GACACCCCG	CGGGCCTGAG	
27661	TGTTGGCCGAT	GTTGGACTTC	ACCGAGCCC	GCAGCACCG	GGTGTGCGC	CCCTCCCCGT	
40	27721	AGGTGGCCAG	CACCGCTGT	GCCTCGATGG	GATGCCAG	CCTGGTGGCG	GTGCCGTGCG
27781	CCTCCACGGC	GTCCACGTCC	GCCGGGGTGA	GCCCAGCGT	GGCCAGGGCC	TGCCGGATCA	
27841	CCCGCTCCTG	CGAGGGCCCG	TTCGGCGCCG	ACAACCGT	GGAAAGCACCG	TCTGGTTGA	
27901	CCGCCGAACC	CCGGACAAAC	GCACGACAC	GGTGGCGT	GGCCTCGGCA	TCGGAGAGCC	
27961	TCTCGACGAT	CAGCACACCG	GACCCCTCGG	CGAAACCGGT	GGCGTCAGCC	GCATCCGCGA	
45	28021	ACGCTTGCA	GCGCGCGTCG	GGCGCGAGAC	CCCGCTGCTG	GGAGAACTCG	ACGAAGCGG
28081	ACGGCGAGGC	CATCACCGT	ACGCCGCCGA	CCAGGGCGAG	CGAGCATTG	CGGAGCGCA	
28141	GTGACTGCC	GGCCTGGTGC	AGCGCCACCA	GCGACGACGA	ACACGCCGTG	TCGACCGTGA	
28201	CCGCCGGACC	CTCCAGACCG	TAGAAGTACG	ACAGCCGACC	GGACAGCACA	CTGGTCTGGG	
28261	TGCCGGTCG	GCCGAAACCG	CCCAGGTCGG	TGCCGAGTC	GTACCCGTCG	GAGAAGGCGC	
50	28321	CCATGAACAC	GGCGGTGTCG	CTTCCCGC	GCGACTCCGG	GAGGATCCC	GGCGTGTCCA
28381	GCGCCTCCCA	CGAGGTCTCC	AGGACCAGAC	GCTGCTGCGG	GTCCATCGCC	AGCGCCCTCAC	
28441	CGCGACTGAT	CCCGAAGAAC	GCCTCGTCGA	AGTCCGCCAC	CCCAGGCGAGG	AAGCCACCAT	
28501	GACGCACGGT	CGACGTGCC	GGATGATCCG	GATCGGGATC	GTACAGCCCG	TCCACGTCCC	
28561	AAACACGGTC	CGTCGAAAC	GCCGTGATCC	CGTCACCA	CGACTCCAGC	ACCCGCCACA	
55	28621	AGTCCTCCGG	CGACCGACCC	CCACCCGGCA	GGCGCAGGC	CATCCCCACG	ATCGCAACAC
28681	GCTCGTCCTG	CGGGACGGCC	GGGGTGTG	TGGGGTGTG	CGATGCCGTG	CGGCGGGACA	
28741	GCGCGCGGT	GAGCTTCGCG	GGACAGCGC	CGGGCGTCCG	GAAGTCGAAG	ACCGGGGTGG	
28801	CGCGCGCG	TACCGCCGTC	GGCTCGGTGA	AGGCGTTGCG	GGCGCGGATC	CCCATGAGCG	
28861	AGTCGACGCC	GAGTCCCTTG	AACTCGCGG	TGCGCTCGAC	CCGTGCGGCA	CGTCCTGGC	
28921	CGAGTACGCC	CGCGGTGAC	TGCCGGACGA	GGCGAGCAC	GTCCCTTTTCG	GGGTCCCGCGG	
60	28981	CGGAGAGCCG	CGCGATCCGG	TGGCGAGGG	TGGTGGCGCC	GGCCGCCCCGG	CGCCGGGGCT

29041 CCCGGCGCGG TCGCGCGAGC AGGGGCGAGC TGCCGAGGCC GGCGGGTCG GCGCGGACCA
 29101 GCGCCGGGTC CGAGGACCGC AACGCCGCGT CGAACACCGT CAGTCGGCCT TCGCGGTCA
 29161 GCGCCGTAC CCGTCGCGG CGCATGCGG CGCGGTGCG GACCGTCAGC CGCTCTCCG
 29221 GTTCCCACAG GCGCCAGGCC ACGGACAACG CGGGCAGTCC CGCTGCCCG CGCTCTTCCG
 5 29281 CCAGCGCTC GAGGAACGCG TTGCGGGCG CGTAGTTCGCG CTGTCCGGG CTGGCGAGCA
 29341 CACCGGGCGC CGACGAGTAG AGGACGAACG CGGCCAGTTC CGTGTCTGG GTGAGTTCT
 29401 GCAGGTGCCA CGCGGGGTCC ACCTTCGGG GCAGCACCGT CTCGAGCCGG TCGGGGGTGA
 29461 GCGCGGTGAG GACGCCGTG TCGAGGACGG CGCGGTGCG CACGACGGCC GTGAGCGGT
 29521 GCGCCGGGTC GATCCCCGCC AGTACGGAGG CGAGTTGCTC CGGGTCGGG ACCTGGCAGG
 10 29581 CGATCGGGT GACCTCGGCC CGGGGCACGT CGCTCGCCGT GCGCTGCGC GACAGCATCA
 29641 GCAGCCGGG CACGCCGTGG CGTTCGACGA GGTGGCGGCT GATGATGCCG GCCAGCGTCC
 29701 CGGAGCCACC GGTGACGAGC ACGGTGCCGT CGGGTGTAG CGCCGGAGCG TCACCCCGCG
 29761 GGACCGCCGG GGCAGACGG CGGGCGTACA CCTGGCCGTC ACGCAGCACC ACCTGGGGT
 15 29821 CATCGAGCGC GGTGGCGCT GCGAGCAGCG GCTCGGGGT CTCCGGGGCG CGTCGACGA
 29881 CGACCGATCCC GCGGGGGTGT TCGGCCCTGCG CGGTCGCGAC CAGTCCGGCG GCGCGGGCG
 29941 ACGCGAGACC GGGCCCGGTG TGGACGGCCA GGACCGCGTC GGCCTACCGG TCGTCGGTGA
 30001 GGAAGCGCTG CACGGCGGTG AGGACGCCGG CGCCCCAGTTC GCGGGTGTGCG TCGAGCGGGG
 30061 CACCGCCGCC GCGTGCGCG GGGAGGATCA CCACGTCCGG GACCGTCGGG TCGTCGAGGC
 20 30121 GGCGGGTCTG CGCGGTCTG GGCAGGACGT CGGGAGCTC GGCAGCACC GGGCGCAGCA
 30181 GGCCCGGAAC GGCTCCCTG ATCGTCAGGG GGCCTCGCG CACGGCGCCG ATGGTGGCGA
 30241 CGGGCCCGCC GGTCTCGTCC GCGAGGTGTA CGCCGTCAGC GGTGACGGCG ACCCGTACCG
 30301 CCGTGGCGCC GGTGGCGTGG ACGCGGACGT CGTCGAACGC GTACGGAAGG TGGTCCCCCTT
 30361 CCGCGCGAG GCGGAGTGC GCGCGAGCA GCGCCGGGTG CAGGCCGTAC CGTCCGGCGT
 30421 CCGCGAGCTG TCCGTGGCG AGGGCCACTT CGGCCAGAC GGCCTCGTGC TCGGCCAGA
 25 30481 CCGCGCGCGG CGGGGCAGC GCGGGCCCGT CGCTGTACCC GGCTCGGGCC AGACGGTCGG
 30541 CGATGTCTC GGGGTCAACC GCGCGGGCGG TGGCGGGCGG CCACGTCGAC GGCATCTCC
 30601 GCACGGCCGG GCGCGCCCG GGGTGGGGGG CGAGGATTCC GTGCGCGTGC TCGGTCACACT
 30661 CCCCCCGCCG GTGCCGCGTG TGCACGGTGA CGCGCGGGCG GCGTCCGCC CGGGCGCGC
 30721 TCACCGTGAC CGAGAGCGCG AGCGCACCGG ACCGGCGCAG CGTGAGGGGG GTGTCCACGG
 30 30781 TGAACGTGTC GAGGGCGCCG CAGCCGGCTT CGTCGCCCGC CGGGATCGCC AGATCCAGGA
 30841 GGGCGCGGC GGGCAGCAC GCGAGGGCGT GCAGGGAGTG CGCCAGCGGA TCGGGCGCGT
 30901 CGACCCCGCC GGTGAGCACC AGGTCGCCGG TGCCGGCAG GGTGACCGCC GCGGTCAGCG
 30961 CGGGGTGCGC GACCGGGTC TGTCCGGCCG GGGCCGCCTC GCGCGGGTC TGGGTGCCGA
 31021 GCCAGTAGCG GACCGCTCG AACGGGTACCG TCGGGGGTG CGAGGGCGT GCGCGCGCG
 35 31081 GGTCGATGAC CTTCGGCCAG TCGACCGTGA CGCCGTCGGT GTGCAGCCGG GCGAGCGCGG
 31141 TCAGGGCGGA TCGCGGTTCG TCGTCGGCGT CGACGATCGG GATGCCGTG ACGAGTCGGG
 31201 TCAGGCTCCG GTCCGGCCG ATCTCCAGGA GCACCGCCCC GTCGTGCGCG GCGACCTGTT
 31261 CCCCCAACCG GACGGTGTG CGGACCTGTC GTACCCAGTA CTCCGGCGTG GTGCAAGGG
 31321 CCCCCCGCCG CTCGGGATC ATCCCGCCCG AGTGGACGC GTGGCTGGTC ACCTTGGGA
 40 31381 ACTCCTCGAG CATCGGCTCC ATCCCGCCCG AGTGGACGC GTGGCTGGTC CGCAGGGCGG
 31441 TGAAGCGGCC GAGCCGGGCC GCGACGTCGA GCACCGCCTC CTCGTACCG GAGAGCACGA
 31501 TCGACGCCGG CCCGTTGACC GCGCGATCT CCACGCCGTC CGCGAGCAGC GGCAGCGCGT
 31561 CCCCCTCCGA CGCGATCACG GCGGCCATCG CCCCCGCCGA CGGCAGCGCC TGCATCAGGC
 31621 GGGCCCGTGC GGACACCAGC CTGACCGCGT CCTCCAGGG CAAGACGCGC GCGACGTACG
 45 31681 CGCGGGCCAG CTCGGGATC GAATGGCCA CGAAGGCCTC CGGGCGTACG CCCCCACGCC
 31741 CGAGCTGTGC CGCGAGTGC ACCTGGAGCG CGAACACCCG GGGCTGGGGG TACCCGGTGT
 31801 CGTGGAGGTC GAGCCCGCG GGCACGTCGA GGGCGCCAG CACCTCGCGG CGAGTGCAGGG
 31861 CGAACAGCTC GTAGGCGCG GCGAGTCCGT CGCCCATGCC GGGACGTTGT GAGCCCTGTC
 31921 CGGAGAAGAG CCACACGAGG CGGCGGTCCG GTTCTCGGGC GCGGGTGACC GTGTCGGTGC
 50 31981 CGATCAGCGC GGCGCGGTGC GGGAGGCCG TGCGGGCGAG CAGGGCCCGCG GCCACCGCGC
 32041 GCTCGTCCTC CTCGGGCGT GCGAGGTGGG CGCGCAGCG GTGTACCTGT CGCTCGASTG
 32101 CCTCGGGGGT CGCGGCCAG AGCAGCAGGG GCAGCGGTCC GGTGTCGGGT CGCGGGCGCG
 32161 GTTCGGGGGC CGGTGGGGG TGGCTTTCGA GGATGATGTG AGCCTGGGTG CGCTAACGC
 32221 CGAAGGAGGA CACCCCGCG CGCCGTGGGC GGTCGTTTC GGGCAGGGG CGGGCGTGG
 55 32281 TGAGGAGTTC GACGGCGCCG CGCGTCCAGT CGACGTGCGA GGACGGCGTG TCCACGTGCA
 32341 GGGTGCAGCG CAGGGCGCC CGGGCGATGG CGAGGACCAT CTTGATGACA CGGGCGACGC
 32401 CCGCGGGCGC CTGAGTGTGG CCGATGTTGG ACTTCGCGA GCGCCAGCAGC ACCGGGGTGT
 32461 CGCGATGCTG CGCGTACCGT GCGAGTACCG CCTGCCCTC GATGGGGTGT CGCGACCTCG
 32521 TCCCGGTGCC ATGCGCTCG ACAGCGTCCA CATCCGCCGG GGTGAGCCCG GCGTTGCCA
 60 32581 GCGCCTGCCG GATCACCCGC TCCTGCGACG GCCCGTTCGG CGCCGACAAC CGCGTTGAAAG

32641 CACCGTCTG GTTGACCGCC GAACCACGCA CGACCGCCAG GACATTGTGG CCGTGCCT
 32701 CGGCGTCGGA GAGCCTCTCG ACGATCAGCA CACCGGATCC CTCGGCGAAA CGGTGCCAT
 32761 CAGCCGATC CGCGAACGCC TTGCAGCGGC CGTCCGGGA GAGGCCCGC TGCTGGGAGA
 32821 AGTCCACGAA GCCGGACGGC GAGGCCATCA CCSTGACGCC CGCGACCAAG CGAGCGAGA
 5 32881 ACTCCCCGAG GCGCAGCGAC TGCCCGGCCT GGTGCAGGCC CACCAAGCGAC GACAAACAC
 32941 CGGTGTCCAC CGTGACCGCC GAGCCCTCA AACCGTAGAA GTACGACAGC CGACCGGACA
 33001 GCACACTGGT CTGGGTGCTG GTGGCACCGA AACCGCCGCC GTCGGCTCA GTGCCGTACC
 33061 CGTAGAAGTA GCCGCCATG AACACGCCGG TGTGCTTCC CGCGAGCGAC TCGGGGAGGA
 33121 TCCCGGCCTG TTCCAGCGCC TCCCACGAGG TCTCCAGGAC CAGACGCTGC TGCGGGTCCA
 10 33181 TCGCCAGCGC CTCACCGCGA CTGATCCCGA AGAACGCCGC GTCGAAGTCC GCCACCCCG
 33241 CGAGGAAGCC ACCATGACGC ACGGTCGACG TGCCCGATG ATCCGGATCG GGATCGTACA
 33301 GCCCGTCCAC GTCCCAACCA CGGTCCGTG GAAACGCCGT GATCCCCTCA CCACCCGACT
 33361 CCAGCAGCCG CCACAAGTCC TCCGGCGACG CGACCCACC CGGCAGCCGG CAGGCCATCC
 15 33421 CCACGATCGC CAACGGCTCG TCCGTCCGGA CGGCCCGGGT CGGGGTACGC CGCCGGGTGG
 33481 TGGCCCGCGC GCCGGCCAGT TCGTCCAGGT GGGCGCGAG CGCCTGCGCC GTGGGGTGGT
 33541 CGAAGACGAG CGTAGCGGC AGCGTCAGGC CGTGCAGCGTC GGCCAGCCGG TTGCGCAGTT
 33601 CGACGCCGGT CAGCGAGTCG AAGCCCACTT CCCTGAACGC CGCGCGGGT GCGATGGCGT
 33661 GGGCGTCGCG GTGGCGGAGC ACCCGCGGAG CGCTGGTACG GACGAGGTG AGCATGTCGC
 20 33721 GCGCGGCCGG AGGTGCGGAC GTGCGCCGGA CGGCCGGCAC GAGGGTGCCT AGGACCGGG
 33781 GGACCCGGTC GGACGGCGCG ACGGCGGCGA GTGCGAGCCG GATCGGACAG AGCGCGGGCC
 33841 GGTGGTGTG CAGGGCCCGC TCGAACAGGG CGAGCCCTG TGCGGCCGTC ATCGGGTCA
 33901 TGCCGTTGG GGCATGCGG GCCAGGTGCG TGGCGGTACG CGCCCGGCC ATCCCGTCCG
 33961 CGCGTCCCA CAGTCCCCAG CGGAGCGAGA CGGCCGGCAG CCCCTGGTGG TGCCGGTGGC
 34021 GGGCGAGCGC GTCGAGGAAC GCGTTGCCGG TCGCGTAGTT GGCGTACCC GCGCCGCCGA
 25 34081 ACATGGCGGA TATGGACGAG TACAGGACGA ACAGGCCAG GTCGAGATCG CGCGTCAGCT
 34141 CGTCAGGTG CCAGGCCAGC TCCGCCTTGA CCCGCAGCAC GGCCTCCAC TGCTCCGGCC
 34201 GCATGGTGTG CACGGCCGCG TCGTCGACGA TCCCGCCAT GTGCACGACG GCGCGCAGCC
 34261 GCTGGCGAC GTCGGCGACG ACTGCGGCCA GTCGTCGCG GTCGACGACG TCGGGGCCA
 34321 CGTACCCGAC GCGGTCTCC TCCGGCGTGT CGCCGGGCCG CCCCTGGCG GACACCACGA
 30 34381 CGACCTCGGC GGCGCTGTG ACGGTGAGCA GGTGGTCCAC GAGGAGGCCG CCGAGCCCG
 34441 CGGTGCCGCC GGTGACGAGG ACGGTCCCAG CGGTCAAGCGG GGAGGTTCCG GTGGCCGCG
 34501 CGACACGGCG CAGACGGGCC GCACCGCTG TGCGTCGGC GACCCGGACG TCGGGCTCGT
 34561 CGCGGGCGC GAGCCGGCC GCTATGGCGG CGGGCGTGT CTCGTCGCT TCGATCAGGG
 34621 CGACCGGGCC GGGATGCTCC GTCTCCGCGC TCCGGACAG GCGCCGAGC GCTTCTCGC
 35 34681 CGGGATCGCC GGTACGGGTG GCCACGATGA GCCGGGATCG CGCCCGAGCG GCGCTGGCGA
 34741 GCCAGGTCTG CACGGTGGTG AGCAGGTGCG GGCCCAGCTC CGGGGTCCCG GCGCCGGCG
 34801 AGGTGCCCGG GTCGCGGGGT TCCACGGCCA GGACACGAC CGGGGGGTGC TCGCGTCGG
 34861 GCACGTCCGC GAGGTACGTC CAGTCGGGA CGGGTACGCG GGGCACGGGC ACCCAGCGA
 34921 TCTCGAACAG CGCCTCGGC TCGGGGTGCG CGGCCCGCAC GGTCAAGCTG TCGACGTCAA
 40 34981 GGACCGGTGA GCGTGTCTG TCCGTGGCA CGATGCGGAC CATGTCGGGG CCGACGCGTT
 35041 CCAGCAGCAC GCGCAGCGCG GTCGCGGCAC GCGCGTGGAT CCTCACGCGC GACCAGGAGA
 35101 ACGCCAGCCG GCGCCGCTCC GGGTCCGTGA AGACCGTCCC GAGGGCGTGC AGGGCCGCGT
 35161 CGAGCAGCAC GGGGTGCGC CCGTACCGGG CGTCGGTGT CTGTTCGGG AGGCGGACCG
 35221 ACGCGTAGGC GCGGCCCTCC CCCGTCCACA TCGCGTCAT GCGCCGGAAC GCGGGCCCGT
 45 35281 ACGAGAGCGG CAGCGCTCG TAGAAGCCGG TCAGGTGCGC CGGGTGGCGC TCGGGGGCG
 35341 GCCAGTCCAC GGGCTCCGCC GGACCGCCAG TGTCCACGCT CAGCGCTCCG GTCGCACTGA
 35401 GCGCCCGAGGG GCGCGTCCG GTACGGCTGT GCAGACTCAC CGACCGCCGT CGGACACCT
 35461 CGGTTCGAC GGTGGCTGG ATCTCCGTGT CGCCGTGCG GTCGACCAACC ACCGGCGCGA
 35521 CGATGGTGTG CTCCCGATC TCCGGCGTGC CGAGCCGGGC TCCCGCTTCG GCGAGCAGTT
 50 35581 CCACGAGCGC CGAGCCGGGC ACGATGACCC GGCGCTCCAC CTCGTTGTCG GCGAGCCAGG
 35641 GCTGACGGCG TACCGAGACA CGCGGGTGGC CAGCGCGCC CGCCCGTCCG GCGAGGTGCA
 35701 CCCACGAGCC GAGCAGCGGG TGGCCGGACG TTCCCGCCCG TTCCGCGTCG ATCCAGTAGC
 35761 GGTACCGGGC GAACGGGTAC GTGGGCAGCG GCACCAACCG ACAGCGTCGCG AACGACCGA
 35821 TGACGGGCAC GCGCCGGAC CAGAGCGCGG CGAGCGACCG AGTGAAGCGG TCCAGGCCGC
 55 35881 CCTCGCTCG CGCGAGTGTG CGGTGACGA CGCTATGCGC ATGCCCGCCG AGCGTGTCT
 35941 CCAGTGCCTG GGTGAGCACG GGATGCGCGC TGACCTGAC GAACGCGCGG TATCCCGGGT
 36001 CGGCCAGGTG CGCGGTGCGC GCGCGAACC GAACGGTGC CGCGAGGTG TCGTACCAAGT
 36061 AGGGCGCGTC CGCGGGCGGG TCCAGGCCACG CCTCGTCCAC GGTGGAGAAG AACGGGACCT
 36121 CGGGCGTGC CGGAGTGTG CGGGCGAGAG CGTCGAGCAG CGCGCCGCCG ATCGTTTCGA
 60 36181 CATGCGCGGT GTGCGACGCG TAGTCGACGG CGATCCGGCG GGCGCGGGGG GTGGCGGCCA

36241 CGAGCTCCTC CACGGCGTCG GCCGCACCGG CGACAAACGAT CGACGCGGGT CCGTTGACCG
 36301 CGCGCACCTC CAGGCGCCCG GCCCACACGG CGGCGTCGAA GTCGGCGGGC GGCAACCGAGA
 36361 CCATGCCGCC CTGCCCGGCC AGTTCGGTGG CGACGAGTCG GCTGCGCACC GCGACGACCT
 36421 CGCGGGCGTC GTCCAGGGTG AGCACCGGGG CGACGCAGGC CGCGGGGACT TCGCCCTGGG
 5 36481 ACTGGCGACG GACCGCGGCC GGGGCGACCC CGTGGCACG CCACAGCTCC GCGAGCGCC
 36541 CCATCACCGC GAACGACGCG GGCTGCACGA CATGCACCG GTCGAACGCG GCGCTCCGG
 36601 CGCGCTGGGC GATGACGTCC AGCAGGTCCC ATCCGGTGTG CGGGGCGAGC GCGTGGCGC
 36661 ACTCGCGGGAG CGCGGGGGCG AACACGGGT CGGTGGCGAG CAGTCGGCA CCCATGCCGG
 36721 CGCACTGGGA GCCCTGCCCG GGGAAACGCGA ACACGACACG TGTGTCGGTG ACGTGGCGG
 10 36781 TTCCCGTCAC GGCCCCCGGC ACTTCGGCAC CACGGCGAA CGCCTCCGCC TCTCGGGCG
 36841 CGACGACCGC CCGGTGGCGC ATGGCCGTC GGGTGGTGGC GAGCGAGTGG CCGACCGCGG
 36901 CGCGGGCGCC AGTGAGCGGG GCCAGCTTC CGCGACGTC CGCAGTCCC TCGGGGGTCC
 36961 CGGCCGACAT CGGCCAGACC ACGTCTCGG GCACCGGCTC GGCTTCGGGT CGGGACACGG
 15 37021 GT3CGGGCGC GGCAGGGGGC CGGGCCTCA GGACGACATG GGCAGTGGTG CGCGTGTGCG
 37081 CGAACGACGA GACACCCGCA CGCCGGGCGC GCCCCGGTAC CGGCCACGGC TCACTGCGG
 37141 GCAGCAGCCG GATGTCGCCG TCCCAGTCCA CGTGGCGGGG CGGCTCGTCG ACGTGAGCG
 37201 TCGCGGGCAG GACGCCGTGC CGCATCGCCA TGACCATCTT GATGACGCCG GCGACGCCGG
 37261 CGCGGGCTG GGTGTGGCCG ATGTTGACT TGAGCGAGCC GATCAGCAGC GGATGACCGC
 20 37321 GTTCGCGCCC GTAGGCCACT TGCAAGGGCT GGGCTCGAC GGGGTGCGG AGACGGGTGC
 37381 CGGTGCCGTG TGCCTCCACG GCGTCGACGT CACCCGGCG CAGGCCGGCG TCGGCGAGCG
 37441 CACGCTGGAT GACGCCGTGC TGCGCAGGCC CGTTGGGGC GGACAGCCCG TTGCGACGCG
 37501 CTCGGAGTT GACCGCGGAG CGCGCACCAC GCGCCAGCAC GGGGTGGCGG TGGCGGGTGG
 37561 CTCGGAGAG CCGCTCCACG ACCAGGACAC CGGCCGCCCTC GCGAAGCTC GTGCCGTCCG
 37621 CGGTGTCCGC GAGGCCCTG GCACGGCCGT CGGGGGCGAG CCCGCGCTGC CGGGAGAACT
 25 37681 CGACGAACCC GGTCGTCGTC GCCATCACCG TGACACCGCC GACCAGGGCG AGCGAGCACT
 37741 CCCCCGAGCG CAGCGACCGC GCGGCCCTGGT GCAGGCCAC CAGCAGCAGC GAACACGCCG
 37801 GTTCGACGGT GACCGACGGG CCCTCCAGAC CGAAGTAGTA CGAGAGCCCG CGGGAGAGAA
 37861 CGCTGGTCGG CGTGGCGTC GCCCCGAAAC CGCCCAGGTC CACGCCCGCG CCGTAGCCCT
 37921 GGGTGAACGC GCCCATGAAT ACGCCGGTGT CGCTGCCCG GACGTTTCG GGCAGGATGC
 30 37981 CGCTCGTTC GAAACGCTCC CACGACGCTT CGAGGACCAAG ACGCTGCTGC GGGTCCATCG
 38041 CGAGCGCCTC ACGCGGGCTG ATCCCCGAAGA ACGCGCGTC GAAAGTCGGCG GCGCCGGTGA
 38101 GGAAGCCGCC GTGACGCACG GAAACCTTGC CGACCGCGTC GGGTTTCGGG TCGTAGAGCG
 38161 CGCGGAGGTC CCAGCCCGGG TCGGCGGGGA ACTCGGTGAT CGCGTCCCCG CGGGAGTCGA
 38221 CGAGCCGCCA CAGGTCTCC GGTGACCGCA CGCCACCGGG CATCCGGCAC GCCATGGCCA
 35 38281 CGATGCCAG CGGCTCGTTC CGCGCACCCG TCGGTGCGGG CACTGTCGCC CGCGAGCGG
 38341 CAGGGGCCG CTCACCCCGC CGTTCCTCAT CGAGCGGGC GGCAGCGCGC GCGGTGTCG
 38401 GTGGTCGAA GACGCCGTG CGGGAGAGCC GTACCCCCGT CGTCTCGGGC AGGCTGTTGC
 38461 GCAACCGGAC ACCGCTGAGC GAGTCGATGC CGAGGTCTT GAAACCGGTC GTGGCGTGA
 38521 TCTCGGAGGC GTGGCGTGG CGGAGCACCG CGGCCGTGGC CGCACACACG ATGGCCAGCA
 40 38581 GCTCACGATC CGGGTCCGG TCGCGGTCS GGTGTCCTC CGCACGGGGC GCGATGCGGC
 38641 GCTCGGTCCG CTGCCGGACG GGCTCGGTGG GAAATGCCGC GACCATGAAC GGCACGTCGG
 38701 CGCGCAGGCC CGCGTCGATG AAGTGGGTGC CCTCGCCCTC GGTGAGCGGC CGGAACCCGT
 38761 CGCGCAGGCC CTGCCGGTC CGTGTGCAA GTTGTCCGGT GAGGGTGCTG GTGGTGTGCC
 38821 ACATGCCCA GGCATGGAG GTGGCGGGTT GGCGAGGGT GTGGCGGTGG GTGGCGAGGG
 45 38881 CTCGAGGAA GGCAGTGGCG GCGCGTAGT TTCTTGTCC GGGGCTGCCG AGGACGGCGG
 38941 CGCGCTGGA GTAGAGGACG AAGTGGGTGA GGGGTGGTT TTGGGTGAGG TGGTGCAGGT
 39001 GCGAGGCCG CTTGGCTTTG GGGTGGAGGA CGGTGGTGAG GCGGTCGGGG GTGAGGGCGT
 39061 CGAGGATGCC GTCGTCGAGG GTGGCGGCCG TGTGAAAGAC GCGGGTGAGG GCTGGGGGGA
 39121 TGTGGCGAG GGTGGTGGCG AGTTGGTGG GGTGCCGCAC GTCGCAAGGGG AGGTGGGTGC
 50 39181 CGGGGGTGGT GTCGGGGGGT GGGGTGCCGG AGAGGAGGT GGTGTGGGG TGTTTCAGGT
 39241 CGGGGGCGAG GATGCCCGCG AGGGTGCCGG AGCCGCCGGT GATGATGATG CGCTGTTCGG
 39301 GTGGTGGGGG GGTGGTGGTG GGTGGGGTGG TGGTCTGGAG GGGGGTGAGG TGGGGTCGGT
 39361 CGAGGGTGTG GTGGGTGAGG CGGAGGTGG GGTGGTCGAG GGTGGCGAGT TGGGCCAGGG
 39421 CGAGGGGAGT GTGGGGGTGG TCGGTTTCA TGAGGCGGAT GCGGTGGGGG TCTTCGTTCT
 55 39481 CGCGGGTGC GGTGAGGCCG GTGACGGTGG CGCCGCCGGG CTCGGTGGTG GTGTGGACGA
 39541 TGAGGGTGTG GTCGGTGGTG GTGAGGTGGT GTTGCAGGGC GGTCAAGACG CGGGTGGCGC
 39601 CGCTGTGGGC CGGGGTGGGT ATGTCCTCGG GGTGTCGGG GTGGCGGGC GTGATCAGGA
 39661 CGCTCCCTC CGGCAGGTCA CGCTCGTAGA CGGCCCTCGGC GACCGCGAGC CACTCCAACG
 39721 CGAGCGGGTT CGGCCCGAC GGGGTGTGG CCCGCTCCCT CAGCACCAGC GAGTCCACCG
 60 39781 ACACGACAGG ACGGCCATCC GGGTCGGCCA CGCGCACGGC GACGCCGGCC TCCCCCGGG

39841 TGAGGGCGAC GCGCACCGCG GCGGCCCGG TGGCGTTAG GCGCACGCC GTCAGGAGA
 39901 ACGGCAGCTC GATCCCGCCG CCCGCGTCGA GCGGCCCGC GTGCAGGGCC CGTCGAGCA
 39961 GTGCCGGATG CACACCGAAA CCGTCCGCCT CGGCGGCCCTG CTCGTCGGC AGGCCACCT
 40021 CGGCATAACAC GGTGTACCA TCACGCCAGG CAGCCCGAA CCCCTGAAAC CGGCACCCCT
 5 40081 ACTCATAACC GGCATCCCGC AGTCGTCAT AGAACCCGA GACGTCGACG CGCGCGGCC
 40141 TGGCCGGCGG CCACTGCGAG AACGGCTCAC CGGAAGCGTT GGAGGTATCC GGGGTGTCG
 40201 GGGTCAGGGT GCCGCTGGCG TGCCGGGTCC AGCTGCCGT GCCCTCGTA CGCGCGTGA
 40261 CGGTACCGG CGGCCGTCCG GCCTCATCGG CCCCTCCAC GGTACCCGAC AGATCCACCG
 40321 CTGCGGTAC CGGCACCCAG AGCGGGGATT CGATGACCAG TTCATCCACC AGCCCGAAC
 10 40381 CGGTCTCGTC ACCGGCCCGG ATGACCAGT CCACAAACGC CGTACCCGGC AGCAGAACCG
 40441 TGCCCCGCAC CGCGTGATCA GCCAGCCAGG GATGCGTACG CAATGAGATC CGGCCGGTGA
 40501 GACAAACACC ACCACCGTCG TCGGCGGGCA GTGCTGTGAC GGCGGCCAGC ATCGGATGCG
 40561 CGGCCCGGT CAGCCCGCC GCGGACAGGT CGTGGCACC GGCGCCCTCC AGCCAGTACC
 40621 GCGCTGTGCTC GAACCGTAG GTGGGCAGAT CGAGCAGCCG CCCCAGCACC GGTGACCA
 15 40681 CGGTGCCCCA GTCCACCCCC GCACCCAGAG TCCACGCCTG CGCCAACGCC CCCAGCCACC
 40741 GCTCCCAGCC ACCGTACCA GTCCGCAACG ACGCCACCGT GCGGGCCTGT TCCATCGCCG
 40801 GCAGCAGCAC CGGATGGCA CTGCACTCCA CGAACACCGA CCCGTCACAGC TCCGCCACCG
 40861 CGGCATCCAG CGCGACAGGG CGACGCAAGGT TCCGGTACCA GTACCCCTCA TCCACCCGGCT
 40921 CGGTACCCCA GGCGCTGTCC ACGGTCGACC ACCACGCCAC CGACCCGGTC CGGCCGGAAA
 20 40981 TCCCTTCAG TACCTCAGCG AGTTGTCCT CGATGCCCTC CACGTGAGGC GTGTGGGAGG
 41041 CGTAGTCGAC CGCGATACGA CGCACCCCGCA CCCCATCAGC CTCATACCAGC GCCACCCACCT
 41101 CCTCCACCGC CGACGGTCC CCCGCCACCA CGTCAAGC CGGACCATTA CGCGCCGCGA
 41161 TCCACACACC CTCGACCAAGA CCCACCTCAC CGGCCGGCAA CGCCACCGAA GCCATCGCCC
 41221 CGCCGCCGGC CAGCCGGCC GCGATCACCC GACTGCGCAA CGCCACCCAGC CGGGCGCGT
 25 41281 CCTCCAGGCT GAGGGCTCCG GCCACACAG CGCCCGCGAT CTCCCCCTGC GAGTGTCCGA
 41341 CCACAGCGTC CGGCACGACC CCATGCGCCT GCCACAGCGC GGCCAGGCTC ACCCGCACCG
 41401 CCCAGCTGGC CGGCTGGACC ACCTCCACCC GCTCCGCCAC ATCCGACCGC GACAACATCT
 41461 CCCGCACATC CCAGCCCGTG TGCGGCAACA ACGCCCGCGC ACACTCCTCC ATACGAGCCG
 41521 CGAACACCGC GGAACGGTCC ATGAGTTCCA CGCCCATGCC CACCCACTGG GCACCCCTGCC
 30 41581 CGGGGAAGAC GAACACCGTA CGCGGCTGAT CCACCGCCAC ACCCATCACC CGGGCATCAC
 41641 CCAGCAGCAC CGCACGGTGA CCGAAGACAG CACGCTCACG CACCAACCCC TGCACGACCG
 41701 CGGCCACATC CACCCCACCC CGCGCAGAT ACCCCCTCCAG CCGCTCCACC TGCCCCCGCA
 41761 GACTCACCTC ACCACGAGCC GACACCGCA ACGGCACCAA CCCATCACCA CCCGACTCCA
 41821 CACCGCAGCG CCCAGGAACA CCCTCCAGGA TCACGTGCGC GTTCGTACCG CTCACCCCGA
 35 41881 ACGACGACAC ACCCGCATGC GGTGCGCGAT CGACTCGGG CCACGGCCTC GCCTCGGTGA
 41941 GCAGCTCCAC CGCACGGGCC GACCAGTCCA CATGCGACGA CGGCTCGTCC ACGTGCAGCG
 42001 TCTTCGGCGC GATCCCATGC CGCATCGCCA TGACCATCTT GATGACACCG GCGACACCCG
 42061 CAGCCGCCTG CGCATGACCG ATGTTGACT TGACCGAAC GAGGTAGAGC GGCGTGTGCG
 42121 GGTCTGTGCC CGTGGCCCG AGGACGGCCT GCGCCTCGAT CGGGTCGCC AGCCCGTGC
 40 42181 CGGTGCGGTG CGCCTCCACC ACCTCCACAT CGGCGCGCG CAGTCGGGGC TTGACCAACG
 42241 CCTGCCGGAT CACCGCGTGC TGGCGACGC CGTTGGGGC GGACAGTCCG TTGGAGGCAC
 42301 CGTCCTGGTT CACCGCCGAG CGCGGACGA CGCGAGAAC GGTGTGCCG TTGCGCTCGG
 42361 CGTGGAGAG CGCTCCAGC ACGAGAACGC CGACGCCCTC GGCGAAGCCG GTCCCGTCCG
 42421 CGCGTCGGC GAACGCCCTG CACCGTCCGT CGGGGGAGAG TCCGCGCTGC CGGGAGAACT
 45 42481 CCACGAGCTC TGCGGTGTT CGCATGACGG TGACACCGCC GACCAGCGCC AGGGAGCACT
 42541 CCCCCGCCCG CAGTGCCTGT CGCGCCTGGT GCAGGGCGAC CAGCGACGAC GAGCACGCCG
 42601 TGTGACCGT GACCGCCGGG CCCTGAAGTC CGTACACGTA CGAGAGGCC CGGGACAGGA
 42661 CGCTCGTCTG CGTCGCCGTG ACACCGAGCC CGCCCGAGTC CGGGCCGACG CGTAGCCCT
 42721 GTTGAACGC GCCCCATGAAC ACGCCGGTGT CGCTCTCCCG GAGCCTGTCC GGCACGATGC
 50 42781 CGGCGTTCTC GAACGCCCTC CAGGAGGTCT CGAGGATCAG GCGCTGCTGG GGGTCCATCG
 42841 CCAGCGCTC GTTCGGACTG ATGCCGAAGA ACGCGCGTC GAACCCGGCG CGGGCCAGGA
 42901 ATCCGCCGTG CGTGTGCGT GAGCGGCCGG CGCGCTCCGG GTCCGGGTG TACAGCGCGT
 42961 CACGTCCCA GCCCCGGTCC GTGGGGAACT CGGTGATCGC CTCGGTACCG GCGCGACGA
 43021 CGGCCACAG GTCTCCGGC GAGGCCACCC CGCCGGCGAG TCGGCACGCC ATGCCGACGA
 55 43081 CGCGACGGG GTCGCCGGAG CGGAGGGTCT GGGCGGTGCG GGGTGCCTCGT GTCGCGGAGC
 43141 CGCGAGGTG CGCGCGAAC GCACCGGGAG TCGGGTGGTC GAACCGGGTT GACGCCGGCA
 43201 CGCGAGACC CGTCCGGCG CGGAGGGTGT TGGTGAACTC GACGGTGGTG AGCGAGTCGA
 43261 CGCGTTCTC CGGAAACGTG CGGTGCGGG AGCAGTGTCC CGCGCCCGGC AGGCCAGGA
 43321 CGGTGGCGAC GCTGTCGGG ACCAGGTGCA GCAGTACGTC CTCCCGGCC GCACGGGCCG
 60 43381 CGCGAGGCG GTTCGCCAC TCCTGTTCCG TGGCGTCGGG CTCGGCCGGT CGGCTCAGTG

43441 CGGTGAGGAT CGGGGGCGTG GCGCCCCCCTA TCGTCGCGGCG CGCGGACCGG
 43501 TCCGGGCCAC GATGTACGAG CGGCCGCCG CGATGGCCTT CTCGATCAGG TCGCCGGTGA
 43561 GCGCCGGCGG TTTCGATGCCG GGCAGCGCGC GGACGGTGAC GGTGGGGAGT CCCTCCGCG
 43621 CCCGTGGCGG GGTGTGGCG TCAGGCGCCGG CGGGGCCGTC GAGCAGGACG TGACGAGCG
 43681 CGCCGGGGTT CGCGGCTTCC TCGGCTGCGG TGGTCACGTG GGTGAGGCCG GTCTCGCG
 43741 GGAGCAGGCC GGCGACGGTG TCAGGCGTCT CCCCGGTGAC CAGGACCGGC GCGTCCGGC
 43801 CGATCGGAGG CGGCACGGTG AGGACCATCT TGCCGGTGTG CCGGGCGTGG CTCATCCAG
 43861 CGAACGCGTC CGGCCACGG CGGATGTCCC ACGGCTGAC CCGCAGCGGG CACAGCTCAC
 43921 CGCGGTGCAA CAGGTGAGG AGCAGTTGA GGATCTCCCG CAGGCGCGC GGATCCACGT
 10 43981 CGGCCAGGTG GAACGGCTGC TGGGCGGCGT GGCGGATGTC GGTCTTGCCT ATCTCGACGA
 44041 ACCGGCCGCC CGGTGCGAGC AGGCCGATGG ACGCGTCGAG GAGTTCACCG GTGAGCGAGT
 44101 TGAGCACGAC GTCGACCGGC GGGAAAGGTG CCGCGAACGC GGCGCTCGG GAGTTCGCCA
 44161 CATGGTCGGT GTCGAAGCCG TCAGGCGTGCAG CAGGTGTTG TTTGGCGGGA CTGGCGGTGG
 44221 CGTACACCTC CGCGCCGAGG TGGCGGGCGA TCCGGGTCGC CGCCATGCCG ACACCGCCCG
 15 44281 TCGCGGCGTG GACCAGGACC TTCTGGCCCG TCGCAGCTC GCGCGCTCG ACCAGGCCGT
 44341 ACCAGGCGGT GGCGAACACG ATGGGCACGG ACGCGGCGAT GGGGAACGAC CATCCCCGTG
 44401 GGATCCGTGC GACCAGCCGC CGGTCCGCGA CCACGCTGCG CCGGAACGCG TCCTGCACGA
 44461 GACCGAACAC CGCGTCGCCG GGGGCCAGGT CGTCGACGCC GGGTCCGACT TCGGTACGA
 44521 TGCCCGCGGC CTCCCCGCCA ATCTCGCCCT CGCCCGGGTA GGTGCCGAGC GCGATCAGCA
 44641 CGCGGGCGGG ACGTCGAGCG GGGCGACGAC GAGGTGCGG AGCGTTCCGG AGGCGGGCGG
 44701 CGCGAGCGGC CACTGCGCG GTCGGCAGGG GGGTGGTGT CGCGCGTAC AGCCGGGGCA
 44761 CGTAGGCCAC GCCGGCCCGC AGCGCGATCT GGGGTTCGCC GAGCGAGGCC GCGGCGGGGA
 44821 CGAGGTCGTC ATCGCCTGCA GTGTCCACCA GCACGAACGA TCCGGGTTCG GCGGCCTGGC
 20 44881 GGGCGAGCGC CTCGCTCCAG AGCCGGGCTG GTTCGCGTC CGGGATCTCG GCGGGGCCGA
 44941 CGCCCCACCGC CGCGGGGGTG ACGACCGTC CGCGGGGTGA CGGGGTGCCG GCGAGGTCGC
 45001 GCCGCTCCCA GACCAGTTG CACAGCGTGG CCTCGCCACT GCGGGTGGCG ACCAGATGGG
 45061 CGCGCAGGCC CGCGAGCCGC CGCGCTGGA CTTGCCCCGA CGCGGTGCGG GGGATCGTGG
 45121 TGACGTGCCA GATCTCGTCG GGCACCTTGA AGTAGGCGAG CGCGCGGCCG CACTCGGCCA
 30 45181 GGATCGCTC CGCGGGGAGC CGGGGGCCGT CGGAAACGAC GTAGAGCACG GGTATGTCGC
 45241 CGAGGACGGG GTGCGGGCGG CGCGCCGCCG CGCGCTCCCG GACACCGGCC ACCTCCCTGGG
 45301 CGACGGTCTC GATCTCCCGG GGGTGGATGT TCTCCCGCC GCGGATGATC AGCTCCTTGA
 45361 CCCGGCCGGT GATCGTCACG TGTCCGGTCT CGGCTGACG TCGCAGGTCC CGGGTGGGT
 45421 ACCAGCCGTC CACGAGCACC TGGCGGGTCG CCTCCGGCTG CGCGTGGTAG CCGAGCATGA
 35 45481 GGCTCGGCCG GCTCGCCAC AGCTCGCCCT CCTCGCCGGG TGCCACGTCG GCGCGGCCA
 45541 CGGGGTCGAC GAACCGCAGC GACAGGCCG GCACGGGCAG CCCGCACGAG CGGGGAACCC
 45601 GCGCATCCTC CAGGGTGTG CGGGTGGAGCG AGCCGGTCGT CTGGTGCAG CCGTACGTGT
 45661 CGAGCAGGGG CACGCCAAC GTCGCCTCGA AATCCCTGGT GAGCGACGCC GCGGAGGTGG
 45721 ATCCGGCGAC CAGCGCCACG CGCAGCGCCG GAGCCCGCG CTCGCGGAC ACGGCGCCGA
 40 15781 GGAGGTTAGCG STACATCGTC GGCACGCCA CGAGCACGGT GCTGGAGTGT TCGGCCAGGG
 45841 CGTCGAGGAC GTCACCGCGC ACGAAGCCGC CCAGGATACG GGCGGACCGC CCGACCGTGA
 45901 GGACGGCGAG CAGGCAGAGG TGGTGGCCGA GGCTGTGGAA CAGCGGGGCC GGCCAGAGCA
 45961 GTTCGTCGTC CTCGGTCAGC CGCCAGGACG GCACGTCGCA GTGCATCGCG GACCACAGGC
 46021 CGCTGCGCTG TGCGGAAACC ACGCCCTTGG GACGGCCGGT GGTGCCGGAG GTGTAGAGCA
 46081 TCCAGGCGGG TTTCGTCAGG CCGAGGTCGT CGCGGGGCCG GCACGGCCGC TCGGCTCCGG
 46141 CGAGGTCCTC STAGGAGACG CAGTCCGGTG CCCGGCGCCC GACGAGCACG ACGGTGGCGT
 46201 CGGTGCCGGT CGGGCGCACC TGGTCGAGGT GGGTTTCGTC GGTGACCGAC ACGGTCGCGC
 46261 CGGAGTCCGT CAGGAAGTGG GCGAGTTCGG CGTCGGCGGC GTCCGGGTG AGCGGGACGG
 46321 CGACGGCGGC GGCGCGGGCG CGGGCGAGGT AGACCTCGAT GGTCTCGATC CGGGTGCAGA
 50 46381 GCAGCATCGC GACCCGGTCG CGCGGGTCGA CGCCGGACGC GCGGAGGTGT CGGGCGAGCC
 46441 GGCGGGCCCG GAGCGGGAGT TCGGTGACG TCACGGCGCG TTGGGAATCC GTGTAGGCGA
 46501 TCCGGTCGCC CGTCGCTCG GCATGGATGC GGAGCAATTG GTGCAACGGC CGGATGGTT
 46561 CCACACGCCG CATGGAAACA CCTTTCTCTC GACCAACCGC ACAACAGCAC GGAACCGGCC
 46621 ACGAGTAGAC GCCGGCGACG CTAGCAGCGT TTTCCGGACC GCCACCCCT GAAGATCCCC
 55 46681 CTACCGTGCC CGGCCCTCCCC GGACGCTCAT CTAGGGGTG GCAAGCATAC CGCCGTGCGT
 46741 AATTGCTTC CTGATGACCG ATGCCGGACG CCAGGGAAAGG GTGGAGGCCGT TGTCCATATC
 46801 TGTACGGCG CGGTATGCC GCTTCGAGAA GACCGGATCA CGCGACCTCG AGGGTGACGA
 46861 GACGGTGTGCG GGCGTGTGAG CGACAGGGCAC CGGCCACACC GACGTGCGC TGGTGGACGG
 46921 TGCTCCCCGG ACCGCCGTGC ACACCCACGAC CCGTACGAC GAGGCCTTCA CCGAGGTCTG
 60 46981 GCACGACAG CGCCCTGTCG AGTCCGGCAT GGACAACGGC ATCGCCTGGG CGGGCACCGA

47041	CGCGTACCTG	TTCGGTGTG	TGCGCACCGG	CGAGAGCGGC	AGGTACGCCG	ATGCCACCGC	
47101	GGCCCTCTAC	ACGAACGTCT	TCCAGCTCAC	CCGGTCGCTG	GGGTATCCCC	TGCTCGCCCG	
47161	GACCTGGAAC	TACGTCAAGCG	GTATCAACAC	GACGAACGCG	GACGGGCTGG	AGGTGTACCC	
47221	GGACTTCTG	GTGGGCCGCG	CCCAGGCCT	CGACGAGGGC	GGGATCGACC	CGGCCACCAT	
5	47281	GCCCGGGGCC	ACCGGTATCG	GCGCCCACCG	GGGGCGGCATC	ACCTGCGTGT	TCCCTGCCGC
47341	CGGGGGCGGA	GTGCGGATCA	ACATCGAGAA	CCCCGCGTC	CTCACGGCCC	ACCACTACCC	
47401	GACGAACGTAC	GGTCCGCGGC	CCCCGGTCTT	CGCACGGGC	ACCTGGCTGG	GCCCGCCGGA	
47461	GGGGGGCGGG	CTGTTCATCT	CCGCGACGGC	CGGCATCCTC	GGACACCGAA	CGGTGCACCA	
10	47521	CGGTGTGTG	ACCGGCCAGT	GCGAGGTGCG	CCTCGACAAAC	ATGGCCCCGGG	TCATCGGC
47581	GGAGAACCTG	CGGCGCCACG	GGGTCCAGCG	GGGGCACGTC	CTCGCCGACG	TGGACCACCT	
47641	CAAGGTCTAC	GTCCGGCGCC	GCGAGGATCT	CGATACGGTC	CGCCGGGTCT	GCGCCGACG	
47701	CCTGTCGAGC	ACCGCGGCCG	TCGCCCTTT	GCACACCGAC	ATAGCCCCCG	AGGATCTGCT	
47761	CGTCGAAATC	GAAGGCATGG	TGGCGTACAA	ATACCCGGTA	AAAGGCCCCG	GACGCTGCGC	
15	47821	CTCGCGGGAT	CCGCGAAGAG	AAAGAAGAGC	GTCACCGCAC	AGCGCGGCAG	CCCGGTCTT
47881	TCGTCCTTCG	CACAGCGCG	GATCTGGTTT	CTCCAGCAAT	TGGACCCGGA	GAGCAACGCC	
47941	TATAATCTCC	CGCTCGTGCA	ACGCGCTCGC	GGTCTATTGG	ACGCGCCGGC	CCTGGAGCGT	
48001	GCGCTGGCGC	TCGTCGTCG	GCGCCACGAG	GGCGTGCAGA	CGGTGTTCGA	CACCGCCGAC	
49061	GGCGAGCCCC	TCCAGCGGGT	GCTTCCCAGCC	CCGGAACACC	TCCTGCGCCA	CGCGCGGGCG	
48121	GGCAGCGAGG	AGGACGCCGC	CCGGCTCGTC	CGCGACGAGA	TGCGCGGCC	TTGACACCTC	
20	48181	GCCACCGGGC	CGTTGATCG	GGCCCTGCTG	ATCCCGCTCG	GTGACGACGA	CCACGTTCTC
48241	GCGGTGACCG	TGCACCATGT	CGCCGGCGAC	GGCTGGTCGT	TGCGGCTCCT	CCACATGAA	
48301	CTCGCAGCCC	ACTACACCGC	GCTGCGCGAC	ACTGCCCCGC	CTGCGAAGT	CCCGCCGTTG	
48361	CCCGGTGAGT	ACGCCGACTT	CGCCGCGCTGG	GAGCGCGCGC	AACTCACCG	CGCCGGACTG	
48421	GACAGCGCTC	TGGCCTACTG	GCGCGAGCAA	CTCCGGGGCG	CCCCGGCGC	GCTCGCCCTC	
25	48481	CCCACCGACC	GTCCCCGGCC	CCGGGTGCGC	GACCGGGACG	CGGGCATGGC	CGAGTGGCGG
48541	CCGCCGGCGC	CGCTGGCCAC	CGCGGTCTCG	ACGCTCGC	GCGACTCCG	TGCGTCCGTG	
48601	TTCATGACCC	TGCTGGCGGC	CTTCCAAGCG	GTCCTCGCC	GGCAGGGGGG	CACCGCGGAC	
48661	GTGCTGGTCG	GCACGGCCGT	GGCGAACCGT	ACGCGGGCGG	CGTACGAGGG	CCTGATCGGC	
48721	ATGTTCGTCA	ACACGCTCG	GCTGCGCGC	GACCTCTCGG	GGGATCCGTC	TTTCCGGGAA	
30	48781	CTCCTCGACC	GCTGCCGGC	CACGACCAACG	GACGCCGTCG	CCCACGCCA	CCTGCCGTT
48841	GAGAACGTCA	TCGAACCTCGT	CCGACCGGAA	CGCGACCTGT	CGGTCAACCC	GGTGTCCAG	
48901	GTGCTGTTGC	AGGTGCTGCG	GCGCGACGCG	GCGACGGCCG	CGCTGCCCG	CATCGCGCC	
48961	GAACCGTTCC	GCACCGGACG	CTGGTTCAC	CGCTTCGACC	TGGAATTCCA	TGTGTACGAG	
49021	GAGCCGGGTG	GCGCGCTGAC	CGCGAACCTG	CTCTACAGCC	GTGCGCTGTT	CGACGAGCCA	
35	49081	CGGATCACGG	GGTTGCTGGA	GGAGTTCACG	GGGGTGCCTC	AGGCGGTAC	CGCCGACCCG
49141	GACGTACGGC	TGTCGCGGCT	GGCGGCCGGC	GACCGACGG	CGGCAGCGCC	CGTGGTGC	
49201	TCGAACGACA	CGGCGCGGG	CCTGCCCCTG	GACACGCTGC	CGGGCTGCT	GGCCCCTGAC	
49261	GCCGCACGCA	CCCCCGGCC	CGTGGCGTC	ACCGACCCGC	ACATCTCCCT	CACCTACGCG	
49321	CAGCTGGACC	GGCGGGCGAA	CCGCCTC3CG	CACCTGCTCC	CGCGCGCGG	CACCGCCACC	
40	49381	GGCGACCTGG	TCGGGATCTG	CGCCGATCGC	GGCGCCGACC	TGATGTCGG	CATCGTGGGG
49441	ATCCTCAAGG	CGGGCGCCGC	TTATGTGCCG	CTGGACCCCC	AACATCCTCC	GGAGCGCAGC	
49501	CGGTTCTGTC	TGGCCGACGC	GCAGCTGACC	ACGGTGGTGG	CGCACGAGGT	CTACCGTT	
49561	CGGTTCCCCG	ATGTGCGCA	CGTGGTGGCG	TTGGACGACC	CGGAGCTGGA	CCGGCAGCCG	
49621	GACGACACGG	CGCCGGACGT	CGAGCTGGAC	CGGGACAGCC	TCGCCCTACGC	GATCTACACG	
45	49681	TCCGGTCTGA	CCGGCAGGCC	GAAGGCCGTG	CTCATGCCG	GTGTCAGCGC	CGTCAACCTG
49741	CTGCTCTGGC	AGGAGCGAC	GATGGGCCGC	GAGCCGGCCA	GCCGCACCGT	CCAGTTCTG	
49801	ACGCCACGT	TCGACTACTC	GGTGCAGGAG	ATCTTTCCG	CGCTGCTGGG	CGGCACGCTC	
49861	GTCATCCCGC	CGGACGAGGT	GCGGTTCGAC	CCGCCCCGGAC	TCGCCCGGGTG	GATGGACGAA	
49921	CAGGCATTAA	CCCGGATCTA	CGCGCCGACG	GCCGTACTGC	CGCGCTGTAT	CGAGCACGTC	
50	49981	GATCCGCACA	GCAGCACAGCT	CGCCGCCCTG	CGGCACCTGT	GCCAGGGCGG	CGAGGCCTG
50041	ATCCTCGACG	CGCGGTTGCG	CGAGCTGTGC	CGGCACCCGGC	CCCACCTGCG	CGTGCACAAT	
50101	CACTACGGTC	CGGCCGAAAG	CCAGCTCATC	ACCGGGTACA	CGCTGCCCGC	CGACCCCGAC	
50161	GCGTGGCCCG	CCACCGCACC	GATCGGCCCG	CCGATCGACA	ACACCCGCAT	CCATCTGCTC	
55	50221	GACGAGGCAGA	TGCGGCCGCG	TCCGGACGGT	ATGCCGGGGC	AGCTCTGCGT	CGCCGGCGTC
50281	GGCCTCGGCC	GTGGGTACCT	GGGCCGTC	GAGCTGACCG	CCGAGCGCTG	GGTGCCTGGGA	
50341	GATGCGGTG	GCGAGGAGCG	CATGTACCTC	ACCGGGCACC	TGGCCCGCCG	CGCGCCCGAC	
50401	GGCGACCTGG	AATTCTCGG	CCGGATCGAC	GACCAAGGTCA	AGATCCGCGG	CATCGCGTC	
50461	GAACCGGGTG	AGATCGAGAG	CCTGCTGCC	GAGGACGCC	GGCTAACGCA	GGCGCGGGTG	
50521	TCCGTGCGCG	AGGACCGGGCG	GGGCGAGAAG	TTCCGGCCG	CGTACGTCGT	ACCGGTGGCC	
60	50581	GGCCGGCACG	GGCACGACTT	CGCCGCCGTC	CTGCGCGCGG	GACTGGCCGC	CCGCGCTGCC

50641 CCCGCGCTCG TGCCCTCCGC CGTCGTCTG GTGGAGCGAC TGCCGAGGAC CACGAGCGGC
 50701 AAGGTGGACC GGCGCGCGT GCCCGACCG GAGGCCGGGC CGGCCTCGAC CGGGCGGGTT
 50761 ACGCCCCGCA CCGATGCCGA CGGGAGGTG TGCCGGATCT TCCAGGAGGT CCTCGACGTC
 50821 CCCCGGGTCG GTGCCGACGA CGACTTCTTG AGCGCTGGCG GGCACCTCCCT GCTCGCCAC
 50881 CCCGGTCTCG CCCGCATCCG CGCCGAGGTG GGTGCCGATG TCCCGCTCGC TACGCTCTTC
 50941 GACGGGGCGGA CGCCCCGCCGC GCTCGCCCT GCAGCGGACG AGGCCGGCCC GGCGCCCTCG
 51001 CCCCGATCG CGCCCTCCGC GGAGAACGGG CGGGCCCCCC TCACCGCGGC ACAGGAACAG
 51061 ATGCTGCACT CGCACGGCTC GCTGCTGCC GCGCCCTCT ACACGGTCGC CCCGTAACGGG
 51121 TTCCGGCTGC CGGGGCCACT CGACC33AA GCGCTCGACG CGGCACGTGAC CGGGATCGCC
 10 51181 CGCGGCCACG AGCCGCTGCG GACCGGGTTC CGCGATCGGG AACAGGTCGT CGGGCCGCC
 51241 GCTCCGGTGC CGCGCGAGGT GGTTCCCTG CGGCTCGGC ACGTCGACGC CGCGGTCGG
 51301 GTCGCCACC GGGAGCTGAC CGGGCCGTT GACCTCGTGA ACGGGTCGTT GCTGCGTGCC
 51361 GTGCTGCTGC CGCTGGCGC CGAGGATCAC GTGCTGCTGC TGATGCTGCA CCACCTCGCC
 51421 GGTGACGGAT GGTCTTCGA CCTCTCTGTC CGGGAGTTGT CGGGGACGCA ACCGGACCTT
 15 51481 CCGGTGTCCT ACACGGACGT GGCCCCGGTGG GAACGGAGTC CGGCCGTGAT CGCGGCCAGG
 51541 GAGAACGACC GGGCTACTG CGGCCGCCGG CTGGGGGGCG CCACCGCGCC GGAGCTGCC
 51601 GCGGTCCGGC CGGGCGGGGC ACCGACC33G CGGGCGTTCC TGTGGACGCT CAAGGACACC
 51661 GCCGTCTCGG CGGCACGCCG GTGCGC33AC GCCCACGACG CGACGTTGCA CGAAACCGTG
 51721 CTCGGCGCCT TCGCCCTGGT CGTGGCGGAG ACCGCCGACA CGACGACGT GCTCGTCGCG
 51781 ACGCCGTTCG CGGACCGGGG GTACGCC33G ACCGACCAAC TCATCGGCTT CTTCGCGAAG
 51841 STCCTCGCGC TGCCTCGCGA CCTCGCGGC ACGCCGTGAT TCCCGAGGT GCTGCGCCGG
 51901 STGCACACCG CGATGGTGGG CGCGCA33CC CACCAAGCGG TGCCCTACTC CGCGCTCGCG
 51961 GCGGAGGACC CGCGCTGCGC GCGGCC33CC GTGCTGTTCC AGCTCATCAG CGCGCTCAGC
 20 52021 CGGAAACTGC GGCTGCCCGG CATGCAC33C GAGCCGTTCC CGTCGTCGCG CGAGACCGTC
 52081 GACGAGATGA CGCGCAACT GTCGATCAC CTCTCGACG ACGGTCGAC CGTCTCCGGC
 52141 CGGGTGGTCC ACGATGCCGC GCTGCTGGC CGTGCACCG TCGACGATT GCTCACCCGG
 52201 STGGAGGCGA CGCTGCGTGC CGCCGCGGGC GACCTCACCG TACGCGTCAC CGGTTACGTG
 52261 GAAAGCGAGT AGCCATGCC GAGCAGGACA AGACAGTCGA TACCTTCGC TGGCGACCCG
 25 52321 CGGAACCTCA GAAGACCCGT GCGGAACCTCG CGCGCACAG CGAGCCGTT GCGATCGTGG
 52381 GGATGGCTG CGGGCTGCC GCGGGGTGCG CGTGCACCG GGACCTGTGG CAGTTGCTGG
 52441 AGTCCGGTGG CGACGGCATC ACCCGCTTCC CCACGGACCG GGGCTGGAG ACCACCGCCG
 52501 ACGGTGCGGG CGGCTTCCTC ACCGGGGCGG CGGCTTCGA CGGGCGTT TCACCGCATCA
 52561 GCCCGCGCGA GCGCTGGCG ATGGACC33C AGCAGCGCT GGCCCTGGAG ACCTCGTGGG
 52621 AGGCCTGCGA GCACGCCGGC ATCGATCCGC AGACGCTGCG GGGCAGTGAC ACGGGGGTGT
 30 52681 TCCCTGGCGC GTTCTTCAG GGGTACGGCA TCGGGCGCGA CTTCGACGGT TACGGCACCA
 52741 CGAGCATTCA CACGAGCGTG CTCTCCGGCC GCCTCGCGTA CTTCTACGGT CTGGAGGGTC
 52801 CGCGGGTCAC GGTGACACCG GCGTGTGGT CGTCGCTGGT GGCGCTGAC CAGGCCGGGC
 52861 AGTCGCTGCG CTCCGGCGAA TGCTCGCTCG CCCTGGTCGG CGCGTCAAG GTGATGGCCT
 52921 CGCCGGCGGG GTTCGCGGAC TTCTCC3AGC AGGGCGGCCT GGCCCCCGAC GCGCGCTGCA
 40 52981 AGGCCTTCGC GGAAGCGGCT GACGGCA33G GTTCGCGCGA GGGGTCCCGC GTCTGATCG
 53041 TCGAGAAGCT CTCCGACGCC GAGCGCA33G GCCACCGCGT GCTGGCGTC GTCCGGGGTT
 53101 CGGCCGTCAA CCAGGACGGT GCCTCC33C GGCTGTCCGC GCCGAACCGGG CGTGCAGCAG
 53161 AGCGGGTGAT CGGGCAGGCC CTGGCC33C CGGGACTCAC CCCGGCGGAC GTGGACGCC
 45 53221 TCGAGGCCA CGGCACCGGC ACCAGGCTGG GCGACCCCAT CGAGGCACAG GCGGTGCTGG
 53281 CCACCTACGG CGAGGGCGC GACACCCCTG TGCTGCTGGG CTGCGTGAAG TCCAACATCG
 53341 GCCACACCCA GGCGCGCGCG GCGTGCACCG GTGTCATCAA GATGGTCCTC GCCATGCGGC
 53401 ACGGCACCCCT GCCCCGCCACC CTGCACGTGG ACACGCCGTC CTCGCACGTC GACTGGACGG
 53461 CGGGCGCCGT CGAACTCTC ACCGACGCC GGCACCGCGT CGAAACCGAC CGCCACCGGC
 53521 GCGCCGGTGT CTCCCTCTC GCGTCA33C GCACCAACGC CCACATCATC CTCGAAAGCC
 50 53581 ACCCCCGACC GGCCCCCGAA CCCGCCCGG CACCCGACAC CGGACCGCTG CCGCTGCTGC
 53641 TCTCGGCCCG CACCCCGCAG GCACTCGACG CACAGGTACA CGCGCTCGGC GCGTCCCTCG
 53701 ACGACAACCC CGGCGCGGAC CGGGTCCCGG TCGCGCAGAC ACTCGCCCGG CGCACCCAGT
 53761 TCGAGCACCG CGCCGTGCTG CTCGGCGACG CGCTCATCAC CGTGAGGCCG AACGCCGGCC
 53821 CGGGACCGGT GGTCTTCGTC TACTCGGGG AAAGCACGCT GCACCCGCAC ACCGGCGGCC
 55 53881 AACTCGCGTC CACCTACCCG GTGTTCGCCG AAGCGTGGCG CGAGGCCCTC GACCAACCTCG
 53941 ACCCCACCCA GGCCCCGGCC ACCGACTTCCG CCCACCGAC CGCGCTCACCG GCGCTCCCTGC
 54001 STCCTGGGG CTCACCCCG CACGCGCTCA CGGGCCACTC CCTCGGTGAG ATCACCGCCG
 54061 CGCACGCCCG CGGTGCTCTG TCCCTGAGGG CGCGGGCGGC GCTCCTCAC ACCCGCACCC
 54121 SCCTGATGGA CCAACTGCCG TCGGGCGGGCG CGATGGTCAC CGTCCTGACC AGCGAGGAAA
 60 54181 AGGCACGCCA GGTGCTGCCG CGGGCGCTGG AGATGCCGC CGTCAACGGC CCCCACCTCC

54241 TCGTGTGTC CGGGGACGAG GAAGCCGTAC TCGAAGCCGC CCGGCAGCTC GGCATCCAC
 54301 ACGGCCTGCC GACCCGCCAC GCCGGCCACT CCGAGCGCAT GCAGCCACTC GTGCGCCCCC
 54361 TCGTCGACGT CGCCCCGGACC CTGACGTACC ACCAGCCCA CACCGCCATC CCCGGCGACC
 54421 CGAACACCGC CGAATACCG GCGCACCCAGG TCGCGCACCA ACTACGTTTJ CAGGGCAC
 5 54481 CGAGCAGTA CCCGGCGCG ACGTTCTCG AGATCGCCCG CAAACAGGAGC CTCTCGCCGC
 54541 TCGTCGACGG CGTTCGGCGC CAGACCGGT ACGCCGCGA GGTGCGGGCG CTGACACACG
 54601 CGTCGCGCA GCTCCACGTC CGCGGCGTGC CGATCGACTG GACGCTCGC CTCGGGGGGG
 54661 ACGCGCGCC CGTCACGCTG CCCACGTATC CGTTCCAGCA CAAGGACTAC TGGCTGCGGC
 54721 CGAACCTCCCG GGCGATGTG ACCGGCGCGG GGCAGGAGCA GGTGGCGCAC CGCTGCTCG
 10 54781 GC3CCCGCGGT CGCGCTGCC GGCACGGGG GAGTCGTCCT GACCGGCCGC CTGTCGCTGG
 54841 CCTCCCACCC GTGGCTCGGC GAGCACGCGG TCGACGGCAC CGTGCCTCTG CCCGGCGCGG
 54901 CCTTCCTCGA ACTCGCGCG GCGGCCGGCG ACGAGGTGCG CTGCGACCTG CTGACGAAC
 54961 TCGTCATCGA GACGCCGCTC GTGCTGCCCG CGACCGCGG TGTGGCGTC TCCGTCGAGA
 55021 TC3CCGAACC CGACGACACG GGGCGCGGG CGGTACCGT CCACGCGGG GCGACCGGCT
 15 55081 CGGGCCTGTG GACCCGACAC GCGCCGCGGAT TCCTCGGCAC GGCACCGGCA CGGCCACCGG
 55141 CCACGGACCC GGCACCCCTGG CGCGCCCGCGG AAGCCGGACC GGTGACGTC GCGACGCT
 55201 ACGACCGGTT CGAGGACATC GGGTACTCCT ACGGACCGGG CTTCCGGGGG CTGCGGGCG
 55261 CCTGGCGCGC CGGCGACACC GTGTACGCG AGGTCGCGCT CCCCGACGAG CAGAGCGCCG
 55321 ACGCCGCCCG TTTCACGCTG CACCCCGCGC TGCTCGACGC CGCGTTCAG CGGCCGCGCG
 20 55381 TGGCCGCGCT CGACGACACC GGCACGGCGG CCCGACTGCC GTTCTCGTTC CAGGACGTC
 55441 GCATCCACCG CGCCGGGGCG ACCGGCGCTGC GGGTCACCGT CGGCGCGAC GCGAGCGCA
 55501 GCACCGCTCG CATGACCGGC CGGGACGGGC AGCTGGTGGC CGTGGTCGGT CGCTGCTCT
 55561 CGCGCCCGTA CGCGGAAGGC TCGGTGACG GCCTCGCTGCG CGCGGTCTGG ACCGAGCTGC
 55621 CGATGCCCGT CGCGTCCGG GACGATCCCG GCGTGGAGGT CCTCGGCGCC GACCCGGCG
 25 55681 ACGCGACGGT TCGGGCGGC ACCCGGGAGC TGACCGCCCG CGTCCTCGGC GCGCTCCAGC
 55741 GCCACCTGTC CGCCGCCGAG GACACCCACCT TGGTGGTACG GACCGGGCAC GGCACGGCG
 55801 CTGGCGCGC CGGGGGTCTG GTCCGCTCGG CGCAGCGGGA GAAACCCGGC CGCGTCGTGC
 55861 TCGTCGAGGC GTCCCCGGAC ACCTCGGTGG AGCTGCTCGC CGCGTGCGCC GCGCTGGACG
 55921 AACCGCAGCT GGGCTCCGG GACGGCGTGC TCTTCGCGCC GCGGCTGGTC CGGATGTCCG
 30 55981 ACCCGCGCA CGGGCCGCTG TCCCTGCCGG ACGGCGACTG GCTGCTCAC CGGTCGCC
 56041 CGGGCACGTT GCACGACGTC GCGCTCATAG CCGACGACAC GCCCCGGCGG GCGCTCGAA
 56101 CGGGCAGGGT CGCGATCGAC GTCCGCGCGG CGGGAACGAA CTTCCGCGAT GTGCTGATCG
 56161 CGCTCGGGAC GTACACCGGG GCCACGGCCA TGGGCGCGA GGGCGCGGGG GTCGTGGTGG
 35 56221 AGACCGGGCC CGGGCTGGAC GACCTGTCCC CGGGCGACCG GGTGTTGGC CTGACCCGGG
 56281 CGGGCATTGG CGCGACGGCC GTCACCGACCG GGCCTGGCT GGCACCGGATC CCCGACGGCT
 56341 GGAGCTTCAC CACGGCGCG TCCGTCCCGA TCGTGGTCGC GACCGCGTGG TACGGCTTGG
 56401 TCGACCTCGG CACACTGCGC GCGGGCGAGA AGGTCTCTGT CAACCGGGCC ACCGGCGTGC
 56461 TCGGCATGCA CGCCGCACAG ATCGCCCGCC ACCTGGCGC CGAGCTCTAC GCCACCGCA
 40 56521 GTACCGGGCA GCAGCACGTC CTGCGCGCC CGGGGCTGCC CGACACGAC ATCGCCGACT
 56581 CTGGACGAC CGCGTTCGG ACCGCTTTCC CGCGCATGGA CGTCGTCCTG AACCGCTGA
 56641 CGGGCGAGTT CTCGACGCG TCGCTCGACC TGCTGGACCC CGACGGCCGG TTCGTCGAGA
 56701 TGGCCGCA CGAGCTGCG GACCCGGCCG CGATCGTCCC CGCCTACCTG CCGTCGACC
 56761 TCGTGGACGC GGGCGCCGAC CGCATCGGG AGATCTGGG CGAACTGCTC CGGCTGTTCG
 56821 ACGCGGGCGC GCTGGAGCCG CTGCGGGTCC GTGCCTGGGA CGTCCGGCAG GACCGCGACG
 45 56881 CGCTCGGCTG GATGAGCCGC GCCCACACA TCGGCAAGAA CGTCCTGACG CTGCCCCGGC
 56941 CGCTCGACCC GGAGGGCGCC GTCGCTCTCA CGGGCGCTC CGGCACGCTC GCGGCATCC
 57001 TCGCCCGCCA CCTGCGCGAA CGCGATGTCT ACCTGCTGTC CGGGACGGCA CGGCCCGAGG
 57061 GGACGCCCGG CGTCCACCTG CCCTCGCGACCG TCGGTGACCG GGACCAAGCTG GCGGGGGCCC
 57121 TGGAGCGGGT GGACCGGGCG ATCACCGCCG TGGTGACCT CGCCGGTGCCT CTGGACGACG
 50 57181 GCACCGTCGC GTCGCTCACCC CGCGAGCGTT TCGACACGGT GCTGCGCCCG AAGGGCGACG
 57241 GCGCCTGGTA CCTGACGAG CTGACGAAGG AGCAGGACCT CGCCCGCTC GTGCTCTACT
 57301 CGTCGGCGCC CGCGCTGCTC GGCAACGCCG GCCAGGGCAA CTACGTCGCC GCGAACGCGT
 57361 TCGTCGACGC GCTCGCCGAG CTGCGCCACCG GTTCCGGCT GCCGGCCCTC TCCATCGGCT
 57421 GGGGGCTCG GAGGGACGTG AGCGGGCTCA CGCGCGCGCT CGGCGAAGCC GACCGGGACC
 55 57481 GGATCGGGCG CAGCGGTTTC CGGGCCATCA CGCGCGAACAA GGGCATGCACT GTGTCAGG
 57541 CGGGCGCCCG CACCGGAAGT CCCGTGGTGG TCGCGCGGC GCTCGACGAC GCGCGGGAGG
 57601 TCGCGCTGCT CGCGGGCTCG CGCGGGACGA CGTCCGGCG GGCCGGCGCTC CGGGAGTGT
 57661 CGTCGGCGCA CGGGCTCGCC CGCGCTGACCG CGGACCGAGCT CGCCGAAGCG CTGCTGACG
 57721 TCGTCGCGGA GAGCACCGCC GCCGTGCTCG GCCACGTGGG TGGCGAGGAC ATCCCCGGCA
 60 57781 CGGGCGCGTT CAAGGACCTC GGCATCGACT CGTCACCGC GGTCCAGCTG CGAACGGCC

57841 TCACCGAGGC GACCGGTGTG CCGCTGAACG CCACGGCGGT CTTCGACTTC CCGACCCCGC
 57901 ACGTGCTCGC CGGAAAGCTC GGCGACGAAC TGACCGGCAC CCGCGCCCGC GTCGTGCCCG
 57961 GGACCGCGGC CACGGCGGT GCGCACGACG AGCCGCTGGC GATCGTGGGA ATGGCCTGCC
 58021 GGCTGCCCGG CGGGGTCGCG TCACCCGAGG AGCTGTGGCA CCTCGTGGCA TCGGGCACCG
 5 58081 ACGCCATCA CGAGATTCCCG ACGGACCGCG GCTGGGACGT CGACGGGATC TACGACCCCG
 58141 ACCCCGACCC GATCGGCAAG ACCTTCGTCC GGACACGGTGG CTTCTCTACC GGCACGACAG
 58201 GCTTCGACCG GGCCTTCTTC GGCACTAGCC CGCCGAGGG CCTCGGATG GACCCGAGC
 58261 AGCGGGTGTG CCTGGAGACG TCGTGGGAGG CGTTGAAAG CGCCGGCATC ACCCCGGACT
 58321 CGACCCGCGG CAGCGACACC GGCGTGTTCG TCGGGCGCTT CCTCTACGGT TACGGCACCG
 10 58381 GTGCGGACAC CGACGGCTTC GGCGCGACCG GCTCGACAGC CAGTGTGCTC TCCGGCCGGC
 58441 TGTCTGACTT CTACGGTCTG GAGGGTCCGG CGGTACCGGT CGACACGGCG TGTTCTCGT
 58501 CGCTGGTGGC GCTGACCAAG GCCGGGCACT CGCTCGCGTC CGGCGAATGC TCGCTGCCCG
 58561 TGGTCGGCGG CGTCACGGTG ATGGCGTCTC CGGGCGGCTT CGTGGAGTTC TCCGGCAGC
 58621 GCGGCCTCGC GCGGACCGC CGGGCGAAGG CGTCGGCGC GGGTGGGAC GGCACGAGCT
 15 58681 TCGCCGAGGG TGCCGGTGTG CTGATCGTC AGAGGCTCTC CGACGCCGAA CGCAACGGTC
 58741 ACACCGTCCT GGCGGTCGTC CGTGGTTCCG CGGTCAACCA GGATGGTGCCT CCAACACGGC
 58801 TGTGGCGCC GAACGGGCCG TCGCAGGAGC GGGTGATCCG GCAGGCCCTG GCAACACGCC
 58861 GGCTCACCCC GGCGGACGTG GACGCCGTG AGGCCACGG CACCGGACCC AGGCTGGCG
 58921 ACCCCATCGA GGCACAGGCG GTACTGGCCA CCTACGGACA GGAGCGGCC ACCCCCCCTGC
 20 58981 TGTCTGGCTC GCTGAAGTCC AACATCGGCC ACGCCACGG CGCGTCCGGC GTCGCCGGCA
 59041 TCATCAAGAT GGTGCAGGCC CTCCGGACG GGGAGCTGCC GCGACGCTG CACGCCGACG
 59101 AGCCGTGCCG GCACGTCGAC TGGACGGCCG GCGCCGTGAA ACTGCTGACG TCGGCCCGGC
 59161 CGTGGCCCGA GACCGACCGG CCACGGCGT CGCGCGTCTC CTCGTTCGGG GTGAGCGGCC
 59221 CCAACGCCCA CGTCATCCTG GAGGCCGGAC CGGTAACCGA GACGCCCGC GCATCGCCTT
 25 59281 CCGGTGACCT TCCCCGTGTG GTGTGGCAC GCTCACCGA AGCGCTCGAC GAGCAGATCC
 59341 GCGCACTGCG CGCCTACCTG GACACCACCC CGGACGTCGA CGGGTGGCC GTGGCACAGA
 59401 CGCTGGCCCG GCGCACACAC TTGCGCCACC GCGCGTGT GCTCGGTGAC ACCCTCATCA
 59461 CCAACACCCC CGCGGACCGG CCGGACGAAC TCGTCTCGT CTACTCCGC CAGGGCACCC
 59521 AGCATCCCGC GATGGCGAG CAGCTCGCG CGGCCCATCC CGTGTTCGCC GACGCCCTGGC
 30 59581 ATGAAAGCGCT CGGCCGCCTT GACAACCCCG ACCCCCCACGA CCCCACGCAC AGCAGCATG
 59641 TGTCTTCGC CCACCAAGCG CGTTCACCG CCCTCTCGC GTCCTGGGC ATACACCCCGC
 59701 ACGCGGTCAT CGGCCACTCG CTGGCGAGA TCACCGGCC GCACGCCGCC GGCATCCTGT
 59761 CGCTGGACGA CGCGTGCACC CTGATCACCA CGCGGCCCG CTCATGCA ACGCTCCGC
 59821 CACCCGGTGC CATGGTCACT GTACTGACCA GCGAAGAGAA GGCAACGCCAG CGGTGGGCC
 35 59881 CGGGCGTGGA GATCGCCGCC GTAACCGGG CCCACTCCAT CGTGTGTCC GGGGACGAGG
 59941 ACGCGGTGCT CACCGTCGCC GGGCAGCTCG GCATCCACCA CGCGCTGCC GCCCCGCACG
 60001 CCGGGCACTC CGCGCACATG GAGCCCGTGG CGCCGAGCT GCTCGCCACC ACCCGCGGGC
 60061 TCCGCTACCA CCCTCCCCAC ACCTCCATTC CGAACGACCC CACCAACGCT GAGTACTGGG
 60121 CCGAGCAGGT CGCGAAGCCC GTGCTGTTC ACGCCACGC GCAGCAGTAC CGGACGCCG
 40 60181 TCTTCGTGGA GATCGCCCCC GCCCAGGACC TCTCCCCCT CGTCGACGGG ATCCCCTGTC
 60241 AGAACGGCAC CGCGGACGAG GTGACGCGC TGACACCGC GCTCGCGCAC CTCTACGCGC
 60301 GCGGTGCCAC GCTCGACTGG CCCCGCATCC TCGGGGCTGG GTCACGGCAC GACCGGGATG
 60361 TCCCCCGCTA CGCGTTCAA CGCGGGCACT ACTGGATCGA GTCGGCACGC CGGCCGCAT
 60421 CCGACGCGGG CCACCCCGTG CTGGGCTCCG GTATGCCCT CGCGGGGTG CGGGGCCGGG
 45 60481 TGTTCACGGG TTCCGTGCCG ACCGGTGCGG ACCGGCGGGT GTTCGTGCC GAGCTGGGCC
 60541 TGGCCGCCGC GGACCGGGTC GACTCGGCCA CGGTCGAGCG GCTCGACATC CCCTCCGTGC
 60601 CCGGGCGGCC GGGCCATGGC CGGACGACCG TACAGACCTG GGTGACGAG CGGGCGGACG
 60661 ACGGCCGGCG CGGGTTCACT GTGCACACCC GCACCGGCCA CGCCCCGTGG ACGCTGCACG
 60721 CCGAGGGGGT GCTGCCCGCC CATGGCACGG CCCTGCCGA TGCGGCCGAC GCCGAGTGGC
 50 60781 CCCCCACCGGG CGCGGTGCCG CGGGACGGGC TGCGGGGTGT GTGGCGCCGG GGGGACCAAGG
 60841 TCTTCGCCGA GGCGAGGTG GACGGACCGG ACGGTTCTGT GGTGCACCCC GACCTGCTCG
 60901 ACGCGGTCTT CTCCGCGGTG GGCGACGGAA GCGCCAGCC GGCGGATGG CGCGACCTGA
 60961 CGGTGCACGC GTCGGACGCC ACCGTACTGC GCGCTGCCT CACCCGGCGC ACCGACGGAG
 61021 CCATGGGATT CGCCGCCTTC GACGGCGCCG GCCTGCCGGT ACTCACCGCG GAGGCGGTGA
 55 61081 CGCTGCGGGG GGTGGCGTCA CGTCCGGCT CGGAGGAGTC GGACGGCTG CACCGGTTGG
 61141 AGTGGCTCGC GGTGGCGAG CGGGTCTACG ACGGTGACCT GCGCAGGGAA CATGCTCTGA
 61201 TCACCGCCG CACCCCGAC GACCCCGAGG ACATACCCAC CGCGCCACAC ACCCGCGCCA
 61261 CCCGGGTCTT GACCCCGCTG CAACACCAC TCACCCACAC CGACCAACACC CTCATCGCC
 61321 ACACCAACAC CGACCCCGCC GGCGCCACCG TCACCCGGCCT CACCCGCACC GCCCAGAACG
 61381 AACACCCCCA CGCGATCCGC CTCATGAAA CGACCAACCC CCACACCCCC CTCCCCCTGG

61441	CCCAACTCGC	CACCCCTCGAC	CACCCCCACC	TCCGCCTCAC	CCACCCACACC	CTCCACCACCC
61501	CCCACCTCAC	CCCCCTCCAC	ACCACCAACC	CACCCACCCAC	CACCCCCCTC	AACCCCCGAAC
61561	ACGCCATCAT	CATCACCGGC	GGCTCCGGCA	CCCTCGCCGG	CATCTCGCC	CGCCACCTGA
61621	ACCACCCCCA	CACCTACCTC	CTCTCCCGCA	CACCCACCCCC	CGACGCCACC	CCCCGCACCC
5	61681	ACCTCCCCCTG	CGACGTCGGC	GACCCCCACC	AACTCGCCAC	CACCCCTCACC
	61741	AACCCCCCTCAC	CGCCATCTTC	CACACCGCCG	CCACCCCTCGA	CGACGGCATC
	61801	TCACCCCCGA	CCGCCTCACC	ACCGTCTCC	ACCCCAAAGC	CAACGCCGCC
	61861	ACCACCTCAC	CCAAAACCAA	CCCCTCACCC	ACTTCGTCT	CTACTCCAGC
10	61921	TCCTCGGGCAG	CCCCGGACAA	GGAAACTACG	CCGGCGCCAA	CGCCTTCCTC
	61981	CCACCCCACCG	CCACACCCCTC	GGCCAACCCG	CCACCTCCAT	CGCCTGGGGC
	62041	CCACCAAGCAC	CCTCACCGGA	CAACTCGACG	ACGGCGACCG	GGACCGCATC
	62101	GTTCCTCTCCC	GATCACGGAC	GACGAGGGCA	TGCGCCTCTA	CGAGGCGGGC
15	62161	GCGAGGACTT	CGTCATGGCC	GCCGCGATGG	ACCCGGCACA	GCCGATGACC
	62221	CGCCCATCCT	GAGCGGCCTG	CGCAGGAGCG	CGCGGCGCGT	CGCCCGTGCC
	62281	TCGCCAGCG	GCTCGCCGAG	CTGCCCAGCG	CCGACCCGCG	CGCGGCCGCTG
	62341	TCTCGGACGC	CACGGCCGCC	GTGCTCGGCC	ACGCCGACGC	CTCCGAGATC
20	62401	CGACGTTCAA	GGACCTCGGC	ATCGACTCGC	TCACCGCGAT	CGAGCTGCGC
	62461	CGGAGGCGAC	CGGGCTGCGG	CTGAGTGCCA	CGCTGGTGT	CGACCACCCG
	62521	TCCTCGCCGC	CAAGCTCCGC	ACCGATCTGT	TCGGCACCGC	CGTCCCCACG
	62581	CGGCACGGAC	CCACCACGAC	GAGCCACTCG	CGATCGTCGG	CATGGCGTGC
25	62641	GCGGGGTGCG	CTCGCCGGAG	GACCTGTGGC	AGCTCGTGGC	GTCCGGCACC
	62701	CCGAGTTCCC	CACCGACCGC	GGCTGGGACA	TCGACCGGCT	TTTCGACCCG
	62761	CCCCCGGCAA	GACCTACGTC	CGGCACCGCG	GCTTCCTCGC	CGAGGCGGCC
	62821	CCGCGTTCTT	CGGCATCAGC	CGCGCGGAGG	CACGGGCCAT	GGACCCGCAG
30	62881	TCTCGAAAC	CTCCTGGGAG	CGTTCGAGA	ACGCGGGCAT	CGTCCGGAC
	62941	GCAGCGACAC	CGGCGTGTTC	ATGGGCGCGT	TCTCCCATGG	GTACGGCGCC
	63001	TGGCGGGGT	CGGCGCCACC	GCCACGCAGA	ACAGCGTGCT	CTCCGGCCGG
	63061	TCTTCGGCAT	GGAGGGCCCG	GGCGTCACCG	TCGACACCCG	CTGCTCGTCG
35	63121	CCCTGCACCA	GGCGGCACAGC	GGCCTCGCGA	CTGGAGAAATG	CTCGCTGGCG
	63181	GTGTACGGT	GATGCCAAC	CCGCTGGGCT	ACGTCGAGTT	CTGCCGCAG
	63241	CCCCCGACGG	CGGTTGCCAG	GCCTTCGCGG	AAGGGCGCCGA	CGGCACGAGC
	63301	GCGCCGGCGT	TCTTGTGCTG	GAGCGGCTCT	CCGACGCCGA	GCGCAACGGA
	63361	TCGCGGTGCGT	CCGCTCTCC	GGCGTCACCC	AGGACGGCGC	CTCCAACGGC
40	63421	CCAACGGCCC	CTCCCACGAG	CGCGTCATCC	GCCAGGGCCCT	CGACAAGGCC
	63481	CCGCGACGT	GGACGTGGTG	GAGGCCACG	GCACCGGAAC	GGGCTCGCCC
	63541	AGGCACAGGC	CATCATCGCG	ACCTACGGCC	AGGACCGCGA	CACACCGCTC
	63601	CGGTCAAGTC	GAACATCGGA	CACACCCAGA	CCACCGCCGG	TGTCGCCCCG
	63661	TGGTCATGGC	GATGCGCCAC	GGCATCGCGC	CGAACACACT	GTACATCAAGA
45	63721	CGCATGTGGA	CTGGACCGAG	GGTGCCTGG	AACTGCTCAC	CGAGGCGAGG
	63781	ACGGGGGACG	CCCGCGCCGC	GCGGGCGTGT	CGTCGCTCGG	TATCAGCGGT
	63841	ACGTGATCCT	TGAGGGTGT	CCCGGGCCCGT	CGCGTGTGGA	GCCGCTGT
	63901	TGCCGTTGCC	GGTGTGGCT	CGGAGTGAGG	CGAGTCTGCG	GGGGCAGGTG
	63961	AGGGGTATCT	CGCGGGGAGT	GTGGATGTGG	CCGGGGCTCGC	GAAGGGGTTG
50	64021	GTGCTGTCTT	CGGTACCCGT	GCGGTACTGC	TGGGTGATGC	CCGGGTGATG
	64081	TGGATCAGCC	CGGTACGGTG	TTCGTCTTC	CCGGGCAGGG	TGCTCAGTGG
	64141	GTGTGGAGTT	GATGGACCGT	TCTGCGGTGT	TCGCGGCTCG	TATGGAGGAG
	64201	CGTTGTTGCC	CCACACGGGC	TGGGATGTGC	GGGAGATGTT	GGCGCGCCCG
	64261	AGCGGGTGGA	GGTGGTCCAG	CCGGCCACGT	GGGCGGTCGC	GGTCAGCTG
	64321	GGCAGGCCCA	CGGGGTCGTA	CCCGACGCCG	TGATCGGACA	GCCGCACTGT
55	64381	CGGCGTGCCT	GGCGGGGGCC	CTCAGCCTTG	AGGACGCCGC	CCGCGTGGTG
	64441	GCCAGGTATC	CGCGGCCGGA	CTGGCCGGGC	GGGGAGCGAT	GGCTTCGGTG
	64501	CGGGTGGAGT	CGGTCTGGTC	GAGGGCGTGT	GGATCGCGC	CGCTAACGGC
	64561	CAGTCGTGGC	CGGCGAGCCG	TGGCGGTGG	AGGACGTGGT	GACGCGGTAT
	64621	GCGTGCAGT	CGTGCAGTAC	GCCGTCGACT	ACGCTCCCA	CACGCCAAC
60	64681	TCGAGGACGA	ACTCGCTGAG	GTACTGAAGG	GAGTGCAGG	GAAGGCCCG
	64741	GGTGGTCGAC	CGTGGACAGC	GCCTGGGTGA	CCGAGCCGGT	GGATGAGAGT
	64801	GGAACCTGCG	TCGCCCCGTC	GGCGCTGGACG	GGCGGGTGGC	GGAGCTGGAC
	64861	TCGTGGAGTG	CAAGCGCCAT	CCGGTGCTCC	TGCCGGCGAT	GGAACAGGCC
	64921	CGTCGTTGCC	CACCGGTGAC	GGCGGGCTGGG	AGCGATGGCT	GACGGCGTTG
	64981	GGACCCCTGGG	CGCGGGCAGTG	GACTGGGACA	CGGTGGTCGA	GGCCGGCTGC

65041	TCGATCTGCC	CACCTACGCG	TTCGAGGCC	GGCGCTACTG	GCTGGAAGCG	GCCGGTGCCA
65101	CCGACCTGTC	CGCGGCCGGG	CTGACAGGGG	CAGCACATCC	CATGCTGCC	GCCATCACGG
65161	CACTACCGC	CGACGACGGT	GGTGTGTC	TCACCGGCC	GATCTCGTT	CGCACGCATC
65221	CCTGGCTGGC	TGATCACGCG	GTGCGGGCA	CGGTCCTGCT	SCCGGGCACG	GCCTTGTGG
5	65281	AGCTGGTCAT	CGGGGCCGGT	GACGAGACCG	GTGCGGGAT	AGTGGATGAA
	65341	AATCCCCCT	CGTGGTGCCG	GCGACCGCAG	CGTGGATCT	CTCGGTGACC
	65401	CTGACGAGGC	CGGACGGCGG	CGAGTGACCG	TCCACGCCG	CACCGAAGGC
	65461	GGACCCGGCA	CGCCAGCGGC	ACCCCTGACCC	CGACACCCCC	CGACACCCCC
	65521	GTGTTGTCGG	TGCGGAGCCG	TTCTCGCA	GGCCACCTGC	CACTGCCCG
10	65581	CCTCCGAGTT	CTACTTGC	CTGGACGCC	TGGGCTACCG	GTTCGGACCC
	65641	GAATGCGGGC	TGCGCTGCC	GATGGTGACA	CCGTGTACGC	CGAGGTGCCG
	65701	ACCGTGC	CGACGCGGAC	GGTTTGGCA	TGCACCCGGC	GCTGCTCGAC
	65761	AGAGCGGCAG	CCTGCTCATG	CTGGAATCGG	ACGGCGAGCA	GAGCGTGCAA
	65821	CCTGGCACCG	CGTCCGGTTC	CACCGCAGGG	GCGCGACCAT	GCTGCGGGTG
15	65881	CGGGCCCGGA	CGGCCTCCGG	CTGCATGCC	CGGACAGCGG	GAACCGTCCC
	65941	TCGACGCGCT	CGTGACCCGG	TCCCCGGAAG	CGGACCTCGC	GCCCACCGAT
	66001	GGGTGGGTG	GGCCCCGGTG	CCCGTACCTG	CCGGGGCCGG	TCCGTCCGAC
	66061	TGACGCTGCG	CGGGCAGCAG	GCCGACCCCG	TCGGGGAGAC	CCGGGACCTG
	66121	TTCTCGACGC	GCTGCTCCGG	GGCGACCGGC	CGGTGATCTT	CCAGGTGACC
20	66181	CCGCCAAGGC	GGCGCAGGC	CTGGTCCGCA	CCGCTCAGAA	CGAGCAGCCC
	66241	TCCTCGTCGA	AACGGACCCG	GGAGAGGTCC	TGGACGGC	GAAGCGCGAC
	66301	CACTCGGC	GCCCCATGTG	CGGCTGCGC	ACGGCCTCTT	CGAGGCAGCC
	66361	GGGCCACGCC	GTCCCTGACG	CTCCCCGGACA	CCGGGTCGTG	CAAGCTGCCG
	66421	CGGTTCCCT	CGACGACCTT	GGCGTGTG	CCACCGACGC	CCCAGACCGG
25	66481	CGGGCAGG	CGGGATCGCG	GTACGCGCG	CGGGCTGAA	CTTCCGGGAT
	66541	C3CTCGGTGT	GGTCGGCGAT	GGCGCTCCGC	TCGGCAGCGA	GGCCGCGGGT
	66601	AGACCGGCC	CGGTGTGAC	GACCTGGCGC	CCGGCAGACCG	GGTCTCTGGG
	66661	GCGCCTTCGG	ACCGGTCGCG	ATCACCGACC	GGCGGCTGCT	CGGCCGGATG
	66721	GGACGTTCCC	GCAGGGCGG	TCCGTGATGA	CCGCGTTCGC	GACCGCGTGG
30	66781	TGACACTGGC	CGGGCTGCC	CCCCGGAGA	AGGTCTGAT	CCACGCGGCC
	66841	TGCGCGGCC	GGCGCTCCAG	ATCGCGCGC	ATCTGGGCGC	GGAGGTGTAC
	66901	GCGCCGCAA	GCGCCATCTG	GTGGACCTGG	ACGGAGCGCA	TCTGGCCGAT
	66961	CCCGGTTCGC	CGACGCGTTC	CCGGCGGTG	ATGTCTGCT	CAACTCGCTC
	67021	TCCTCGACGC	GTCCGTCGGC	CTGCTCGCG	CGGGTGGCCG	GTTCATCGAG
35	67081	CGGACATCCG	GCACGCCGTC	CAGCAGCCG	TCGACCTGAT	GGACGCCGCG
	67141	TGCAGCGGAT	CATCGTCGAG	CTGCTCGGC	TGTCGCGCG	CGACGTGCTG
	67201	CGGTCCACGC	CTGGGACGTG	CGGCAGGCC	GGGAGGC	CGGCTGGATG
	67261	GTACACACCG	CAAGCTGGT	CTGACGGTCC	CGCGGCCG	GGATCCCCGAG
	67321	TCATCACCGG	CGGCTCCGGC	ACCCCTGCC	GCATCCTCGC	CCGCCACCTG
40	67381	ACACCTACCT	GCTCTCCCGC	ACCCACCCC	CCGACACCA	CCCCGGCAC
	67441	GCGACGTCG	CGACCCCCAC	CAACTCGCA	CCACCCCTCG	CAACCCCTCA
	67501	CCGGCGTCTT	CCACACCGCC	GGAACCCCTG	ACGACGCC	GTCGACAAAC
	67561	ACCGCGTCG	CACCGCTCTC	AAACCCAAGG	CCGACGCC	CTGGCACCTG
	67621	CCCGCGACAC	CGACCTCGCC	GCGTCGTC	TCTACTCCG	GGTCGCCG
45	67681	GCCCCGGCA	GGGCAACTAC	GTCGCGCGA	ACGCGTCT	CGACGCGCTC
	67741	CCCGTGC	AGGGCTGCC	GCGCAGTCC	TCGCATGGG	CATGTGGCG
	67801	CGCTCACCGC	GAAGACTACC	GACCGGGACC	GCCAGCGCAT	CCGGCGCAGC
	67861	CGTTGAGCGC	CGCGGACGGC	ATGCGGCTG	TCGACGCC	GACCGACCGG
	67921	TCGTGTCG	GACGACCGTC	GACCTACCC	AGCTGACGG	CCCGTGC
50	67981	CGGGCTGGC	CGCGCACCGG	GGCGGCC	CGCGCACGG	CGCCCGCAAC
	68041	AGCCCCCTGGC	CGTGC	GGCGCGTA	CGCGGCC	CGAGCGGCC
	68101	AGGTGTCGCT	CCGCCACGCG	GGCGCGGTC	TCGCGTACGG	GCTGGCGAC
	68161	CGGACCGTCC	GGTCCGCGAG	CTCGGTTTCG	ATTGCGTAC	CGCGGTGAC
	68221	GGCTCGCGGC	CGAGACGGG	CTGCGGCTG	CGACGACGCT	GGTGTTCAGC
55	68281	CGGAGGCGCT	CACCGCCAC	CTGCTCGAC	TGATCGACG	TCCCACCGCC
	68341	GGGAGTCCCT	GGCCGCGGTG	ACGGCGCTC	CCGTGGCG	CCGCGGGGAC
	68401	CGATCGCCAT	CGTGGCGATG	GGCGTCCG	TGCGGCG	TGTGACGTCG
	68461	TGTGGCGGCT	CGTCGAGTCC	GGCACCGACG	CGATCACCAC	GCCTCTGAC
	68521	GGGACGTCGA	CGCGCTGTAC	GACCGGGACC	CGGACGCC	CGGCAAGGCG
60	68581	GGGGCGGT	CCTGGCCGG	GGGGCGGAGT	TCGACGCC	TACACCTGC

68641	GCGAAGCGCT	CGGCATGGAC	CCGCAGCAAC	GCCTGCTGCT	CGAAACGGCG	TGGGAGGCGA	
68701	TCGAGCGCGG	CCGGATCAGT	CCGGCGTCGC	TCCGCGGCCG	GGAGGTCGGC	GTCTATGTCG	
68761	GTGCGGCCGC	GCAGGGCTAC	GGGCTGGCG	CCGAGGACAC	CGAGGGCCAC	GCGATCACCG	
68821	GTGGTTCCAC	GAGCCTGCTG	TCCGGACGGC	TGGCGTACGT	GCTCGGGCTG	GAGGGCGCCG	
5	68881	CGGTCAACCGT	GGACACGGCG	TGCTCGTCGT	CTCTGGTCGC	GCTGCATCTG	GCGTGCAGG
68941	GGCTGCGCCT	GGGCGAGTGC	GAACTCGCTC	TGGCCGGAGG	GGTCTCCGTA	CTGAGTTGCG	
69001	CGGCCGCGTT	CGTGGAGTTC	TCCCAGCAGC	CGGGCGTCGC	GGCCGACGGG	CGCTGCAAGT	
69061	CGTCGGCGC	GGGCGCGGAC	GCGACGACGT	GGTCCGAGGG	CGTGGCGTGT	CTCGTACTGT	
10	69121	AACGGCTCTC	CGACGCCAGG	CGGCTGGGGC	ACACCGTGT	CGCCGTCGTC	CGGGCAGCG
69181	CCGTCAACGTC	CGACGGCGCC	TCCAACGGCC	TCACCCCGCC	GAACGGGCTC	TCGAGCAGC	
69241	GGGTCATCCG	GAAGGGCGTC	GGCCGGGCCG	GGCTGACCGG	CGCCGACGTG	GACGTCGTG	
69301	AGGGGCACGG	CACCGGCACC	CGGCTCGGGC	ACCCGGTCSA	GGCGGACCGC	CTGCTCGCA	
69361	CGTACGGCA	GGACCGTCCG	GCACCGGTCT	GGCTGGGCTC	GCTGAAGTCG	AACATCGAC	
15	69421	ATGCCACGGC	CGCGGCCGGT	GTCGCGGGCG	TCATCAAGAT	GGTGCAGGGC	ATCGGCGCG
69481	GCACGATGCC	CGGGACGCTG	CATGTGGAGG	AGCCCTCSCC	CGCCGTCGAC	TGGAGCACCG	
69541	GACAGGTGTC	CCTGCTCGGC	TCCAACCGGC	CCTGGCCCGA	CGACGAGCGT	CGCGCCCGGG	
69601	CGGCCGTCTC	CGCGTTCGGG	CTCAGCGGGA	CGAACCGCGA	CGTCATCTG	GAACAGCACC	
69661	GTCCGGCGCC	CGTGGCGTCC	CAGCCGCCCG	GGCCGCCCCG	TGAGGAGTCC	CAGCCGCTGC	
69721	CGTGGGTGCT	CTCCCGCGGG	ACTCCGGCCG	CGCTGCGGGC	CAAGGCGGCC	CGGCTGCGCG	
20	69781	ACCACCTCGC	GGCGGCACCG	GACGCGGATC	CGTTGACAT	GGGGTACCGG	CTGGCCACCA
69841	GCCCCGCCCA	GTTCGCCCAC	CGTGCCGCGG	TCGTGCCAC	CACCCCGGAC	GGATTCCGTG	
69901	CCGCGCTCGA	CGGCCTCGCG	GACGGCGCGG	AGGCGCCCGG	AGTCGTACCC	GGGACCGCTC	
69961	AGGAGCGGGC	CGTCGCTTC	CTCTTCGACG	GCCAGGGCGC	CAAGCGCGCC	GGAATGGGGC	
70021	CGGAGCTCCA	CCGCCGTTT	CCCGTCTTCG	CCGCCGCGTG	GAACGAGGTC	TCCGACCGT	
25	70081	TCGGCAAGCA	CCTCAAGCAC	TCCCCCACGG	ACGTCATCCA	CGCGAACAC	GGCGCTCTCG
70141	CCCATGACAC	CCTGTACGCC	CAGGCCGGCC	TGTCACGCT	CGAAGTGGCG	CTGCTCGCG	
70201	TGCTGGAGCA	CTGGGGGGTG	CGGCCGGACG	TGCTCGTCGG	GAACCTCCGTC	GGCGAGGTGA	
70261	CCGGCGCGTA	CGCGGCGGGG	GTGCTCACCC	TGGCGACGC	GACGGAGTTG	ATCGTGGCCC	
30	70321	GGGGCGGGGC	GCTCGGGGGC	CTGGCGCCCC	GGCGATGCT	CGCCGTCGAC	GGAAGCCCGG
70381	CGGAGGTGG	CGCCCGCACG	GATCTGGACA	TCGCCCGGT	CAACGGCCCG	TCCGCCGTGG	
70441	TGCTGCCCGG	TTCGCCGGAC	GATGTGGCGG	CGTTCGAACG	GGAGTGGTCC	CGGGCCGGGC	
70501	GGCGCACGAA	ACGGCTCGAC	GTCGGGCACG	CGTTCCACTC	CGGGCACGTC	GACGGTGC	
70561	TCGACGGCTT	CCGTACGGTG	CTGGAGTCGG	TCGGCTTCGG	CGGGGGCGGG	CTGCCGTGG	
35	70621	TGTCCACGAC	GACGGGCCGG	GACGCCGCGG	ACGACCTCAT	AACGCCCGCG	CACTGGCTGC
70681	GCCATGCGCG	TCGGCCGGTG	CTGTTCTCGG	ATGCCGTCCG	GGAGCTGGCC	GACCGCGCG	
70741	TCACCACGTT	CGTGGCCGTC	GGCCCCCTCGG	GCTCCCTGGC	GTCGGCCGCG	GCGGAGAGCG	
70801	CCGGGGAGGA	CGCCGGGACC	TACCAACGCGG	TGCTGCGCG	CGGGACCGGT	GAGGAGACCG	
70861	CGGGCGTGCAC	CGCCCTCGCC	GAGCTGCACG	CCCACGGCGT	CCCGGTCGAC	CTGGCCGCGG	
40	70921	TACTGGCCGG	TGGCCGGCCA	GTGGACCTTC	CCGTGTACGC	GTTCCAGCAC	CGTTCTACT
70981	GGCTGGCCCC	GGCCGTGGCG	GGGGCGCCGG	CCACCGTGGC	GGACACCGGG	GGTCCGGCG	
71041	AGTCCGAGCC	GGAGGACCTC	ACCGTCGCCG	AGATCGTCCG	TCGGCGCAC	CGGGCGCTGC	
71101	TCGGCGTCAC	GGACCCCGCC	GACGTCGATG	CGGAAGCGAC	GTTCTCGCG	CTCGGTTTCG	
71161	ACTCACTGGC	GGTGCAGCGG	CTGCGCAACC	AGCTCCCTC	GGCAACCGGG	CTGGACCTGC	
45	71221	CGGCGGCCGT	CCTGTTCGAC	CACGACACCC	CGGCGCGCT	CACCGCGTTC	CTCCAGGACC
71281	GGATCGAGGC	CGGCCAGGAC	CGGATCGAGG	CGGGCGAGGA	CGACGACCGG	CCCACCGTGC	
71341	TCTCGCTCCT	GGAGGAGATG	GAGTCGCTG	ACGCCGCGGA	CATCGGGCG	ACGCCGGCCC	
71401	CGGAGCGTGC	GGCCATCGCC	GATCTGCTG	ACAAGCTCGC	CCATACCTGG	AAGGACTACC	
71461	GATGAGCACC	GATACGCAACG	AGGGAACGCC	GGCCGCCGGC	CGCTGCCCAT	TCGCGATCCA	
50	71521	GGACGGTCAC	CGGCCATCC	TGGAGAGCGG	CACGGTGGGT	TCGTCGACC	TGTTGGCGT
71581	CAAGCACTGG	CTGGTCGGCG	CCGGCGAGGA	CGTCAAGCTG	GTCACCAACG	ATCCCGGGTT	
71641	CAGCTCGGCC	GCGCCCTCCG	AGATGCTGCC	CGACCGGGG	CCCGGCTGGT	TCTCCGGGAT	
71701	GGACTCACCG	GAGCACAACC	GCTACCGGCA	GAAGATCCG	GGGGACTTCA	CACTGCGCCC	
71761	GGCCCGCAAG	CGGGAGGACT	TCGTCGCCGA	GGCCGCCGAC	GCCTGCCCTGG	ACGACATCGA	
71821	GGCCCGGGGA	CCCGGCACCG	ACCTCATCCC	CGGGTACGCC	AAGCGGCTGC	CCTCCCTCGT	
55	71881	CATCAACGCG	CTGTACGGC	TCACCCCTGA	GGAGGGGCC	GTGCTGGAGG	CACGGATGCG
71941	CGACATCACC	GGCTCGGCCG	ATCTGGACAG	CGTCAAGACG	CTGACCGACG	ACTTCTTCGG	
72001	GCACCGCGCTG	CGGCTGGTCC	GGCGAAGCG	TGACGAGCGG	GGCGAGGACC	TGCTGCAACCG	
72061	GCTGGCCTCG	GCCGACGACG	GCGAGATCTC	GTCAGCGAC	GACGAGGCGA	CGGGCGTGT	
72121	CGCGACGCTG	CTGTTGCCG	GCCACGACTC	GGTGCAGCAG	ATGGTCGGCT	ACTGCCTCTA	
60	72181	CGCACTGCTC	AGCCACCCCG	AGCAGCAGGC	GGCGCTGCGC	GCGCGCCCGG	AGCTGGTCCG

72241 CAACGCGGTC GAGGGAGATGC TCCGTTTCTC GCCCGTCAAC CAGATGGCG TACCGCGCGT
 72301 CTGTGTCGAG GACGTCGATG TCGGGGGCGT GCGCATCCGT GCGGGCGACA ACGTGATCCC
 72361 GCTCTACTCG ACGGCCAACCG GCGACCCCGA GGTGTTCCCG CAGCCCGACA CCTTCGATGT
 72421 GACGCGCCCG CTGGAGGGCA ACTTCGCGT CCGCCACGGC ATTACAAGT GTCCCGGCCA
 5 72481 GCACATCGCC CGGGTGTCA TCAAGGTCGC CTGCGTGC GGTTGTCGAGC GTTCCCGGA
 72541 CGTCCGGCTG GCGGGCGACG TGCCGATGAA CGAGGGCTC GGGCTGTTCA GCGGGGCCGA
 72601 GCTGCGGTC ACCTGGGGG CGGCATGAGT CACCCGGTGG AGACGTTGCG GTTGCACGAC
 72661 GGGACGACGG TCGCGCACAT CAACGCGGGC GAGGCGCAGT CCTCTTACCG GGAGATCTC
 72721 ACCCAGCGCT GCTACCTGCG CCACGGTGT GACCTGC CGGGGGACGT GGTGTTGAC
 10 72781 GTCGGCGCGA ACATCGGCAT GTTACGCTT TTGCGCGATC TGGAGTGTCC TGGTGTGACC
 72841 GTGCACGCC TCGAGGCCGC GCGCGTGC CGGGCTGCGC TGCGGGCGAA CGTACGCGG
 72901 CACGGCATCC CGGGCCAGGC GGACCAGTGC GCGGTCTCG ACAGCTCCGG CACCCGGAAAG
 72961 ATGACCTTCT ATCCCGACGC CACGCTGATG TCCGGTTTC ACACGGATGC CGCGGCCCG
 15 73021 ACGGAGCTGT TGCGCACGCT CGGGCTCAAC GGCAGCTACA CGCGCGAGGA CGTCGACACC
 73081 ATGCTCGCGC AACTGCCGA CGTCAGCGAG GAGATCGAAA CCCCTGTGGT CGGCTCTCC
 73141 GACGTATCG CGGAGCGCGG TATCGAGGCC ATCGGCCTGC TGAAGGTCGA CGTGGAGAAG
 73201 AGCGAACCGC AGGTCTTCGC CGGCCTCGAG GACACCGACT GGCCCCGTAT CGGCCAGGTC
 73261 GTCGCGGAGG TCCACGACAT CGACGGCGCG CTCGAGGAGG TCGTCACCGCT GCTCCGCGC
 20 73321 CATGGCTTCA CCGTGGTCGC CGAGCAGGAA CGCTGTTCG CGGGCACGGG CATCCACAG
 73381 GTCGCCGCGC GGCGGGTGGC CGGCTGAGCG CGTCGCGGGC CGCGGCCGTC CGCACCGGC
 73441 GCGCGGGTGC GGACGGCGGC TCAGCCGGCG TCGGACAGTT CCTTGGCAG TTGTCGACGG
 73501 CCCTTCACCC CCAGCTGCG GAACACGTT GTGAGGTGCT GTTCCACCGT GCTGGAGGTG
 73561 ACGAACAGCT GGCTGGCGAT CTCTTGTG TGCGGCCCGA CGCGGGCGTG CGACGCCACC
 73621 CGCGCTCCG CCTCGGTCA CGATGTGATC CGCTGCGCCG CGTCACGTC CTGGTGC
 25 73681 TCCGCGTCCG AGGACTCCCC ACCGAGCCGC CGGAGGAGCG GCACGGCTCC GCACTGGTC
 73741 GCGAGGTGCC GTGCGCGCG GAACAGTCCC CGCGCACGGC TGTGCGCGCG GAGCATGCCG
 73801 CACGCTTCGC CCATGTCGGC GAGGACGCGG GCGAGCTCGT ACTGGTCGCG GCACATGATG
 73861 AGCAGATCGG CGGCCTCGTC GAGCAGTTG ATCCGTTGG CCGCGGACT GTAGGCCGCC
 73921 TGCACCCGCA GCGTCATCAC CGCGGCCCGG GACCCATCG GCCGGGACAG CTGCTCGGAG
 30 73981 ATGAGCCTCA GCCCCCTCGTC ACGGCCCGGG CGAGCAGCA GAAGCGCTTC GGCAGCGTC
 74041 ACCCGCACCA GGGCCAGGCC CGGCACGTC ACGGACCGAG GTCGCATCCG CTCCCGCAG
 74101 TCCCCGAACG CGTTGTACGC CGCCCGGTAC CGCCCGGCCG CGAGATGGTG TTGCCACGG
 74161 GCGCAGACCA TGTGCACTGC GAAGAGGCTG TCGGAGGTCT CCTCCGGCAA CGGCTCGGCG
 74221 AGCCACCGCT CGGCCCGGT CAGGTCGCC CGTCGAGTC CGGGGGCCAC GGTGCTGTC
 35 74281 AGCGGCAATG CGCGGGCAT CCCCCAGGAG GGCACGACCC GGGGGCGAG CGCGGCCCTCG
 74341 CCGCATTGCA CGCGGGCGGT CAGGTCGCC CGGCAGCGC CGGCCTCGGC CGGAAACCC
 74401 GCGTGGACCG CCTCGTCGGC CGGGTCCGC ATGTTGTCGT CACCGGCCAG CTTGTCGACC
 74461 CAGGACTTGG ACGCATTGGT GTCCTCGGC TAGAGCAGGG CCAGCAACGC CATCATGGTC
 74521 GTGGTCCGGT CGTCGTGAC CGGGAGTGC TGGAGCACGT ACTCGGCTT GGCCTGGCC
 40 74581 GTTCGGACC AGCCGCGCAG CGCGTTGCTC AGGGCTTGT CGCGACGGC GCGGTGCC
 74641 ACGGCTCCGG AAAACGAGGC GACCTCGTC TCGGCCGGCG GATGGCCCG ACAGGGCGGA
 74701 TCGGCCGCGC CGGGATAGAT CAGCGCGAGG GACAGGTCCG CGACGCGCAG GTGCGCC
 74761 CCCTGCTCGC TCGGGCGGC GGAGCGCTGG CGCGCCAGGA CCTCGCGCGC CTCGCCCC
 74821 CGCCCGTCCA TCGCCAGCCA CGAGGCGAGC GACACGGCGT GCTCGCTGGA GAGGAGCG
 45 74881 TCCCCGACG CGGTGAGCAG CTCGGGCACA TGCCGGCCGG ATCTGGCGGG ATCGCAGAGC
 74941 CGCTCGATGG CGCGGGTGT GACGCGCAGT CGGGCGTGG CGCGGGGGTC GTCGGAGGCC
 75001 CGGTAGGCGA ACTCCAGGTA GGTGACGGCC TCGTCAGCT CGCCGCGCAG GTGGTGC
 75061 CGCGCGCGT CGGTGAACAG CCCGGCGACC TCGGCGCCGT GACACCGGGC GGTACCCATC
 75121 TGGTGGCGGG CGAGCACCTT GCTGGCCACG CGCGGGTCCC GCAAGCTTC CAGGCCAGC
 50 75181 TCGTGCAGGC CACGCCGCTC GGGCGGGAG AGGTCGTCGA GTACGACGGA CGGGCGCG
 75241 GGGTGCAGGA ACCGCCCTTC CGCGAGCAGC CGCCCGCTCGA CGAGCTGTC GTGGCGCTG
 75301 TCGACCGCCT CGGTGTCGAG GCCGGTCATC CGCTGGACGA GGGTGAGTTC GACACTCTG
 75361 CCGAGCACGG CGGAAGCTCG CGCGACGCTC AGCGGGCCCG CGCCGCAACG ATAGAGCGAC
 75421 CCGAGGTAGG CGAGCCGGTA CGCCCGCCCC GCGACCACTT CGAGGACACC TGAGGTCC
 55 75481 GTCCGTGCCT CGCGATGTC GTGATCAGG CGTGGCCGA GGAGCAGGTT CGCGCGCG
 75541 GCGCGAACG CCTGGGCCAC CACGTCGTC TGCGCGTCT CGCGAGGTG CGCGCGCAGC
 75601 AGTTCGGTGG TCTCGCTC GGTGAGCGGG CGCAGCGCA TCTCTGGTA GTGGCGCAGA
 75661 CTCAGCAGTG CGGCCGGAA TTGGGAGTGG CGGGCGTGC GCGGGAGCAG CTCGGTCAGC
 75721 ACGATGGCGA CACGGGCCCG GCTGATGCGG CGCGCGAGGT GGAGCAGGCA GCGCAGCGAC
 60 75781 GGCGCGTCGG CGTGGTCAC GTGTCGATG CCGATCAGTA CGGGCCGCTC CGCGCGCAGC

75841	GTCAGCACCG	TGCGGGTGAG	TCGGTCCCC	AGGCGGTTGT	CGACGTCGGC	CGGCAGGTTT	
75901	TCGCACGATG	CCGTCAGCCG	GACCAGCTCC	GGTGTCCGGG	CGGCCAGCTC	GGGCTGGTCG	
75961	AGGAGCTGGC	CGAGCATGCC	GTACGGCAGG	GCCCCTCCT	CCATGGAGCA	CACCGCGCGA	
76021	AGGGTACCGA	AGCGGGCTT	GGCGCGGGCG	GCCTCGAGGA	GTTCGGTCTT	GCGCGAGGCG	
5	76081	ATCGGCCCGG	TGACGGCGGC	GACGACGCC	CGCCCGCCCC	CCGCTCGGGT	GAGCGCCCGG
76141	TGGAGGGAAC	CGAACTCGTC	ATCGCGGGCG	ATCAGGTCTG	GGGGAGATAA	GCGCGCTATC	
76201	ACGAATGGAA	CTACCTCGCG	ACCGTCGTGG	AAACCCATAG	GCATCACATG	GCTTGTGAT	
76261	CTGTACGGCT	GTGATTCA	CTGGCGGGAT	GCTGTGCTAC	AGATGGGAAG	ATGTGATCTA	
76321	GGGCCGTGCC	CTTCCCTCAG	GAGCCGACCG	CCCCCGGCC	CACCCGCCGT	ACCCCTGGG	
10	76381	CCACCAAGCTC	GGCGACCCGC	TCCTGGTGGT	CGACGAGGT	GAAGTGCCCG	CGGGGAAAGA
76441	CCTCCACCGT	GGTCGGCGCG	GTCTGTGCC	CGGCCCAGGC	GTGGGCTCTG	TCCACCGTCG	
76501	TCTTCGGATC	GTCTCACCG	ATGCACACCG	TGATCGCGT	CTCCAGCGGC	GGCGCGGGCT	
15	76561	CCCACCGGTA	CGTCTCCGCC	GGCTAGTAGT	CCGCCCGCAA	CGCGCCAGG	ATCAGCGCG
76621	GCATTTCGTC	GTCCGCCATC	ACATCGGCC	TCGTCGCC	GAGGCCGATG	ACCGCCGCCA	
76681	GCAGCTCGTC	GTCTGACGCG	AGGTGGTCT	GGTCGGCGCG	CGGCTGCGAC	GGCGCCCGCC	
76741	GGCCCGAGAC	GATCAGGTGC	GCCACCGGG	GCCGCTGGGC	CAGCTGAAC	GCGAGTGTG	
20	76801	CGCCCCATGCT	GTGGCCGAAC	AGCACCA	GACGGTCCAG	CCCCGGCTTC	AACGCCCTGG
76861	CCACGAGGCC	GGCGAGAAC	CGCAGGTGC	GCACCGCCTC	CTCGTCGCC	CGGTCTGGC	
76921	GGCCGGGTTA	CTGCACGGCG	TACACGTC	CCACCGGGG	GAGCGCACGG	GCCAGCGGAA	
25	76981	GGTAGAACGT	CGCCGATCCG	CCGGCGTGGG	GCAGCAGCAC	CACCCGTACC	GGGGCCTCGG
77041	GGCTGGGAA	GAACGCGCG	AGCCAGAGTT	CCGAGCTCAC	CGCACCCCC	CGGCCGCGAC	
77101	CTGGGGAGCC	CGGAACCGGG	TGATCTCGC	CAAGTGCTTC	TCCCGCATCT	CGGGGTCGGT	
77161	CACGCCCAT	CCCTCCTCCG	GCGCCAGACA	GAGGACGCG	ACTTTGCGT	TGTGCACATT	
77221	GCGATGCACA	TCGCGCACCG	CCGACCCGAC	GTCGTCGAGC	GGGTAGGTCA	CCGACAGCGT	
25	77281	CGGGTGCACC	ATCCCCCTGC	AGATCAGGGC	GTTCCCTC	CACGCCCTCAC	GATAGTTCG
77341	GAAGTGGGTA	CCGATGATCC	GTTCACCGA	CATCCACAGG	TACCGATTGT	CAAAGGCGTG	
77401	CTCGTATCCC	GAGGTTGACG	CGCAGGTGAC	GATCGTGCCA	CCCCGACGTG	TCACGTAGAC	
77461	ACTCGCGCCG	AACGTCGCGC	GCCCCGGGTG	CTCGAACACG	ATGTCGGGAT	CGTCACCGCC	
30	77521	GGTCAGCTCC	CGGATC				

Those of skill in the art will recognize that, due to the degenerate nature of the genetic code, a variety of DNA compounds differing in their nucleotide sequences can be used to encode a given amino acid sequence of the invention. The native DNA sequence encoding the FK-520 PKS of *Streptomyces hygroscopicus* is shown herein merely to illustrate a preferred embodiment of the invention, and the present invention includes DNA compounds of any sequence that encode the amino acid sequences of the polypeptides and proteins of the invention. In similar fashion, a polypeptide can typically tolerate one or more amino acid substitutions, deletions, and insertions in its amino acid sequence without loss or significant loss of a desired activity. The present invention includes such polypeptides with alternate amino acid sequences, and the amino acid sequences shown merely illustrate preferred embodiments of the invention.

The recombinant nucleic acids, proteins, and peptides of the invention are many and diverse. To facilitate an understanding of the invention and the diverse compounds and methods provided thereby, the following general description of the FK-520 PKS genes and modules of the PKS proteins encoded thereby is provided. This general description is followed by a more detailed description of the various domains and modules of the FK-520

PKS contained in and encoded by the compounds of the invention. In this description, reference to a heterologous PKS refers to any PKS other than the FK-520 PKS. Unless otherwise indicated, reference to a PKS includes reference to a portion of a PKS. Moreover, reference to a domain, module, or PKS includes reference to the nucleic acids encoding the same and vice-versa, because the methods and reagents of the invention provide or enable one to prepare proteins and the nucleic acids that encode them.

The FK-520 PKS is composed of three proteins encoded by three genes designated *fkbA*, *fkbB*, and *fkbC*. The *fkbA* ORF encodes extender modules 7 - 10 of the PKS. The *fkbB* ORF encodes the loading module (the CoA ligase) and extender modules 1 - 4 of the PKS. The *fkbC* ORF encodes extender modules 5 - 6 of the PKS. The *fkbP* ORF encodes the NRPS that attaches the pipecolic acid and cyclizes the FK-520 polyketide.

The loading module of the FK-520 PKS includes a CoA ligase, an ER domain, and an ACP domain. The starter building block or unit for FK-520 is believed to be a dihydroxycyclohexene carboxylic acid, which is derived from shikimate. The recombinant DNA compounds of the invention that encode the loading module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of methods and in a variety of compounds. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for the loading module of the heterologous PKS is replaced by the coding sequence for the FK-520 loading module, provides a novel PKS coding sequence. Examples of heterologous PKS coding sequences include the rapamycin, FK-506, rifamycin, and avermectin PKS coding sequences. In another embodiment, a DNA compound comprising a sequence that encodes the FK-520 loading module is inserted into a DNA compound that comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the loading module coding sequence is utilized in conjunction with a heterologous coding sequence. In this embodiment, the invention provides, for example, either replacing the CoA ligase with a different CoA ligase, deleting the ER, or replacing the ER with a different ER. In addition, or alternatively, the ACP can be replaced by another ACP. In similar fashion, the corresponding domains in another loading or extender module can be replaced by one or more domains of the FK-520 PKS. The resulting heterologous loading module coding sequence can be utilized in conjunction

with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide.

The first extender module of the FK-520 PKS includes a KS domain, an AT domain specific for methylmalonyl CoA, a DH domain, a KR domain, and an ACP domain. The 5 recombinant DNA compounds of the invention that encode the first extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 first extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for 10 a module of the heterologous PKS is either replaced by that for the first extender module of the FK-520 PKS or the latter is merely added to coding sequences for modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the first extender module of the FK-520 PKS is inserted into a DNA compound that comprises the remainder of the coding sequence 15 for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or only a portion of the first extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2- 20 hydroxymalonyl CoA specific AT; deleting either the DH or KR or both; replacing the DH or KR or both with another DH or KR; and/or inserting an ER. In replacing or inserting KR, DH, and ER domains, it is often beneficial to replace the existing KR, DH, and ER domains with the complete set of domains desired from another module. Thus, if one desires to insert an ER domain, one may simply replace the existing KR and DH domains with a KR, DH, 25 and ER set of domains from a module containing such domains. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a gene for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting 30 heterologous first extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous

PKS can be replaced by one or more domains of the first extender module of the FK-520 PKS.

In an illustrative embodiment of this aspect of the invention, the invention provides recombinant PKSs and recombinant DNA compounds and vectors that encode such PKSs in which the KS domain of the first extender module has been inactivated. Such constructs are especially useful when placed in translational reading frame with the remaining modules and domains of an FK-520 or FK-520 derivative PKS. The utility of these constructs is that host cells expressing, or cell free extracts containing, the PKS encoded thereby can be fed or supplied with N-acylcysteamine thioesters of novel precursor molecules to prepare FK-520 derivatives. See U.S. patent application Serial No. 60/117,384, filed 27 Jan. 1999, and PCT patent publication Nos. US97/02358 and US99/03986, each of which is incorporated herein by reference.

The second extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the second extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 second extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the second extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the second extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the second extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these

- replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous second extender module coding sequence
- 5 can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the second extender module of the FK-520 PKS.

The third extender module of the FK-520 PKS includes a KS, an AT specific for

10 malonyl CoA, a KR, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the third extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 third extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS.

15 The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the third extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the third extender module of the FK-520 PKS is inserted into a DNA

20 compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, all or a portion of the third extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the

25 malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting the KR and/or the inactive DH; replacing the KR with another KR; and/or inserting an active DH or an active DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding

30 sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous third extender module coding sequence can be utilized in conjunction with a coding sequence from a PKS that synthesizes FK-520, an

FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the third extender module of the FK-520 PKS.

- The fourth extender module of the FK-520 PKS includes a KS, an AT that binds
- 5 ethylmalonyl CoA, an inactive DH, and an ACP. The recombinant DNA compounds of the invention that encode the fourth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fourth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS.
- 10 The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the fourth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the fourth extender module of the FK-520 PKS is inserted into a
- 15 DNA compound that comprises the remainder of the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the fourth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the ethylmalonyl CoA

20 specific AT with a malonyl CoA, methylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; and/or deleting the inactive DH; inserting a KR, a KR and an active DH, or a KR, an active DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of

25 the FK-520 PKS, a PKS for a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous fourth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fourth extender

30 module of the FK-520 PKS.

As illustrative examples, the present invention provides recombinant genes, vectors, and host cells that result from the conversion of the FK-506 PKS to an FK-520 PKS and vice-versa. In one embodiment, the invention provides a recombinant set of FK-506 PKS

genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by those for the AT domain of the fourth extender module of the FK-520 PKS. This recombinant PKS can be used to produce FK-520 in recombinant host cells. In another embodiment, the invention 5 provides a recombinant set of FK-520 PKS genes but in which the coding sequences for the fourth extender module or at least those for the AT domain in the fourth extender module have been replaced by those for the AT domain of the fourth extender module of the FK-506 PKS. This recombinant PKS can be used to produce FK-506 in recombinant host cells.

Other examples of hybrid PKS enzymes of the invention include those in which the 10 AT domain of module 4 has been replaced with a malonyl specific AT domain to provide a PKS that produces 21-desethyl-FK520 or with a methylmalonyl specific AT domain to provide a PKS that produces 21-desethyl-21-methyl-FK520. Another hybrid PKS of the invention is prepared by replacing the AT and inactive KR domain of FK-520 extender module 4 with a methylmalonyl specific AT and an active KR domain, such as, for 15 example, from module 2 of the DEBS or oleandolide PKS enzymes, to produce 21-desethyl-21-methyl-22-desoxo-22-hydroxy-FK520. The compounds produced by these hybrid PKS enzymes are neurotrophins.

The fifth extender module of the FK-520 PKS includes a KS, an AT that binds methylmalonyl CoA, a DH, a KR, and an ACP. The recombinant DNA compounds of the 20 invention that encode the fifth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 fifth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS 25 is either replaced by that for the fifth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS. In another embodiment, a DNA compound comprising a sequence that encodes the fifth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the FK-520 PKS or a recombinant FK-520 PKS that 30 produces an FK-520 derivative.

In another embodiment, a portion of the fifth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA

- specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one or both of the DH and KR; replacing any one or both of the DH and KR with either a KR and/or DH; and/or inserting an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions,
- 5 the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous fifth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another
- 10 polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the fifth extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the DH domain of the fifth extender module have been deleted or mutated to render the DH non-functional. In one such mutated gene, the KR and DH coding sequences are replaced with those encoding only a KR domain from another PKS gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-19 to C-20 double bond of FK-520 and has a C-20 hydroxyl group. Such analogs are preferred neurotrophins, because they

15 have little or no immunosuppressant activity. This recombinant fifth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this fifth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth

20 extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (lacking the C-19 to C-20 double bond of FK-506 and having a C-20 hydroxyl group) FK-506 derivative. In another embodiment, the present invention provides

25 a recombinant FK-506 PKS in which the DH domain of module 5 has been deleted or otherwise rendered inactive and thus produces this novel polyketide.

30

The sixth extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds

of the invention that encode the sixth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 sixth extender module is inserted into a DNA compound that comprises the coding sequence for a 5 heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the sixth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the sixth extender module of the FK-520 PKS is 10 inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the sixth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA 15 specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific AT; deleting any one, two, or all three of the KR, DH, and ER; and/or replacing any one, two, or all three of the KR, DH, and ER with another KR, DH, and ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a 20 coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous sixth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous 25 PKS can be replaced by one or more domains of the sixth extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the DH and ER domains of the sixth extender module have been deleted or mutated to render them non-functional. In one such 30 mutated gene, the KR, ER, and DH coding sequences are replaced with those encoding only a KR domain from another PKS gene. This can also be accomplished by simply replacing the coding sequences for extender module six with those for an extender module having a methylmalonyl specific AT and only a KR domain from a heterologous PKS gene, such as.

for example, the coding sequences for extender module two encoded by the *eryA1* gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that has a C-18 hydroxyl group. Such analogs are preferred neurotrophins, because they have little or no immunosuppressant activity. This recombinant sixth extender module 5 coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this sixth extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides 10 recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (having a C-18 hydroxyl group) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the DH and ER domains of module 6 have been deleted or otherwise rendered inactive and thus produces 15 this novel polyketide.

The seventh extender module of the FK-520 PKS includes a KS, an AT specific for 2-hydroxymalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds of the invention that encode the seventh extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. 20 In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 seventh extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the seventh extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the 25 heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the seventh extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

30 In another embodiment, a portion or all of the seventh extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 2-hydroxymalonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or

malonyl CoA specific AT; deleting the KR, the DH, and/or the ER; and/or replacing the KR, DH, and/or ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous seventh extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the seventh extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant FK-520 PKS genes in which the coding sequences for the AT domain of the seventh extender module has been replaced with those encoding an AT domain for malonyl, methylmalonyl, or ethylmalonyl CoA from another PKS gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-15 methoxy group, having instead a hydrogen, methyl, or ethyl group at that position, respectively. Such analogs are preferred, because they are more slowly metabolized than FK-520. This recombinant seventh extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this seventh extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (C-15-desmethoxy) FK-506 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the AT domain of module 7 has been replaced and thus produces this novel polyketide.

In another illustrative embodiment, the present invention provides a hybrid PKS in which the AT and KR domains of module 7 of the FK-520 PKS are replaced by a methylmalonyl specific AT domain and an inactive KR domain, such as, for example, the AT and KR domains of extender module 6 of the rapamycin PKS. The resulting hybrid PKS produces 15-desmethoxy-15-methyl-16-oxo-FK-520, a neurotrophin compound.

The eighth extender module of the FK-520 PKS includes a KS, an AT specific for 2-hydroxymalonyl CoA, a KR, and an ACP. The recombinant DNA compounds of the invention that encode the eighth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a 5 DNA compound comprising a sequence that encodes the FK-520 eighth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either replaced by that for the eighth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a 10 novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the eighth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the eighth extender module coding sequence is 15 utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the 2-hydroxymalonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or malonyl CoA specific AT; deleting or replacing the KR; and/or inserting a DH or a DH and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these 20 replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous eighth extender module coding sequence can be utilized in conjunction with a PKS that synthesizes FK-520, an FK-520 derivative, or 25 another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the eighth extender module of the FK-520 PKS.

In an illustrative embodiment, the present invention provides a set of recombinant 30 FK-520 PKS genes in which the coding sequences for the AT domain of the eighth extender module has been replaced with those encoding an AT domain for malonyl, methylmalonyl, or ethylmalonyl CoA from another PKS gene. The resulting PKS genes code for the expression of an FK-520 PKS that produces an FK-520 analog that lacks the C-13 methoxy group, having instead a hydrogen, methyl, or ethyl group at that position, respectively. Such

analog are preferred, because they are more slowly metabolized than FK-520. This recombinant eighth extender module coding sequence can be combined with other coding sequences to make additional compounds of the invention. In an illustrative embodiment, the present invention provides a recombinant FK-520 PKS that contains both this eighth 5 extender module and the recombinant fourth extender module described above that comprises the coding sequence for the fourth extender module AT domain of the FK-506 PKS. The invention also provides recombinant host cells derived from FK-506 producing host cells that have been mutated to prevent production of FK-506 but that express this recombinant PKS and so synthesize the corresponding (C-13-desmethoxy) FK-506 10 derivative. In another embodiment, the present invention provides a recombinant FK-506 PKS in which the AT domain of module 8 has been replaced and thus produces this novel polyketide.

The ninth extender module of the FK-520 PKS includes a KS, an AT specific for methylmalonyl CoA, a KR, a DH, an ER, and an ACP. The recombinant DNA compounds 15 of the invention that encode the ninth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 ninth extender module is inserted into a DNA compound that comprises the coding sequence for a module of a heterologous PKS. The resulting construct, in which the coding sequence for a module of 20 the heterologous PKS is either replaced by that for the ninth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the ninth extender module of the FK-520 PKS is 25 inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion of the ninth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the methylmalonyl CoA specific AT with a malonyl CoA, ethylmalonyl CoA, or 2-hydroxymalonyl CoA specific 30 AT; deleting any one, two, or all three of the KR, DH, and ER; and/or replacing any one, two, or all three of the KR, DH, and ER with another KR, DH, and/or ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can

originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or from chemical synthesis. The resulting heterologous ninth extender module coding sequence can be utilized in conjunction with a PKS that synthesizes FK-520, an FK-520 derivative, or 5 another polyketide. In similar fashion, the corresponding domains in a module of a heterologous PKS can be replaced by one or more domains of the ninth extender module of the FK-520 PKS.

The tenth extender module of the FK-520 PKS includes a KS, an AT specific for malonyl CoA, and an ACP. The recombinant DNA compounds of the invention that encode 10 the tenth extender module of the FK-520 PKS and the corresponding polypeptides encoded thereby are useful for a variety of applications. In one embodiment, a DNA compound comprising a sequence that encodes the FK-520 tenth extender module is inserted into a DNA compound that comprises the coding sequence for a heterologous PKS. The resulting construct, in which the coding sequence for a module of the heterologous PKS is either 15 replaced by that for the tenth extender module of the FK-520 PKS or the latter is merely added to coding sequences for the modules of the heterologous PKS, provides a novel PKS coding sequence. In another embodiment, a DNA compound comprising a sequence that encodes the tenth extender module of the FK-520 PKS is inserted into a DNA compound that comprises the coding sequence for the remainder of the FK-520 PKS or a recombinant 20 FK-520 PKS that produces an FK-520 derivative.

In another embodiment, a portion or all of the tenth extender module coding sequence is utilized in conjunction with other PKS coding sequences to create a hybrid module. In this embodiment, the invention provides, for example, either replacing the malonyl CoA specific AT with a methylmalonyl CoA, ethylmalonyl CoA, or 2- 25 hydroxymalonyl CoA specific AT; and/or inserting a KR, a KR and DH, or a KR, DH, and an ER. In addition, the KS and/or ACP can be replaced with another KS and/or ACP. In each of these replacements or insertions, the heterologous KS, AT, DH, KR, ER, or ACP coding sequence can originate from a coding sequence for another module of the FK-520 PKS, from a coding sequence for a PKS that produces a polyketide other than FK-520, or 30 from chemical synthesis. The resulting heterologous tenth extender module coding sequence can be utilized in conjunction with a coding sequence for a PKS that synthesizes FK-520, an FK-520 derivative, or another polyketide. In similar fashion, the corresponding domains in a

module of a heterologous PKS can be replaced by one or more domains of the tenth extender module of the FK-520 PKS.

The FK-520 polyketide precursor produced by the action of the tenth extender module of the PKS is then attached to pipecolic acid and cyclized to form FK-520. The 5 enzyme FkbP is the NRPS like enzyme that catalyzes these reactions. FkbP also includes a thioesterase activity that cleaves the nascent FK-520 polyketide from the NRPS. The present invention provides recombinant DNA compounds that encode the *fkbP* gene and so provides recombinant methods for expressing the *fkbP* gene product in recombinant host cells. The recombinant *fkbP* genes of the invention include those in which the coding 10 sequence for the adenylation domain has been mutated or replaced with coding sequences from other NRPS like enzymes so that the resulting recombinant FkbP incorporates a moiety other than pipecolic acid. For the construction of host cells that do not naturally produce pipecolic acid, the present invention provides recombinant DNA compounds that express the enzymes that catalyze at least some of the biosynthesis of pipecolic acid (see 15 Nielsen *et al.*, 1991, *Biochem. 30*: 5789-96). The *fkbL* gene encodes a homolog of RapL, a lysine cyclodeaminase responsible in part for producing the pipecolate unit added to the end of the polyketide chain. The *fkbB* and *fkbL* recombinant genes of the invention can be used in heterologous hosts to produce compounds such as FK-520 or, in conjunction with other PKS or NRPS genes, to produce known or novel polyketides and non-ribosomal peptides. 20 The present invention also provides recombinant DNA compounds that encode the P450 oxidase and methyltransferase genes involved in the biosynthesis of FK-520. Figure 2 shows the various sites on the FK-520 polyketide core structure at which these enzymes act. By providing these genes in recombinant form, the present invention provides recombinant host cells that can produce FK-520. This is accomplished by introducing the recombinant 25 PKS, P450 oxidase, and methyltransferase genes into a heterologous host cell. In a preferred embodiment, the heterologous host cell is *Streptomyces coelicolor* CH999 or *Streptomyces lividans* K4-114, as described in U.S. Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference. In addition, by providing recombinant host cells 30 that express only a subset of these genes, the present invention provides methods for making FK-520 precursor compounds not readily obtainable by other means.

In a related aspect, the present invention provides recombinant DNA compounds and vectors that are useful in generating, by homologous recombination, recombinant host

- cells that produce FK-520 precursor compounds. In this aspect of the invention, a native host cell that produces FK-520 is transformed with a vector (such as an SCP2* derived vector for *Streptomyces* host cells) that encodes one or more disrupted genes (i.e., a hydroxylase, a methyltransferase, or both) or merely flanking regions from those genes.
- 5 When the vector integrates by homologous recombination, the native, functional gene is deleted or replaced by the non-functional recombinant gene, and the resulting host cell thus produces an FK-520 precursor. Such host cells can also be complemented by introduction of a modified form of the deleted or mutated non-functional gene to produce a novel compound.
- 10 In one important embodiment, the present invention provides a hybrid PKS and the corresponding recombinant DNA compounds that encode those hybrid PKS enzymes. For purposes of the present invention a hybrid PKS is a recombinant PKS that comprises all or part of one or more modules and thioesterase/cyclase domain of a first PKS and all or part of one or more modules, loading module, and thioesterase/cyclase domain of a second PKS.
- 15 In one preferred embodiment, the first PKS is all or part of the FK-520 PKS, and the second PKS is only a portion or all of a non-FK-520 PKS.
- One example of the preferred embodiment is an FK-520 PKS in which the AT domain of module 8, which specifies a hydroxymalonyl CoA and from which the C-13 methoxy group of FK-520 is derived, is replaced by an AT domain that specifies a malonyl, 20 methylmalonyl, or ethylmalonyl CoA. Examples of such replacement AT domains include the AT domains from modules 3, 12, and 13 of the rapamycin PKS and from modules 1 and 2 of the erythromycin PKS. Such replacements, conducted at the level of the gene for the PKS, are illustrated in the examples below. Another illustrative example of such a hybrid PKS includes an FK-520 PKS in which the natural loading module has been replaced 25 with a loading module of another PKS. Another example of such a hybrid PKS is an FK-520 PKS in which the AT domain of module three is replaced with an AT domain that binds methylmalonyl CoA.
- In another preferred embodiment, the first PKS is most but not all of a non-FK-520 PKS, and the second PKS is only a portion or all of the FK-520 PKS. An illustrative 30 example of such a hybrid PKS includes an erythromycin PKS in which an AT specific for methylmalonyl CoA is replaced with an AT from the FK-520 PKS specific for malonyl CoA.

- Those of skill in the art will recognize that all or part of either the first or second PKS in a hybrid PKS of the invention need not be isolated from a naturally occurring source. For example, only a small portion of an AT domain determines its specificity. See U.S. provisional patent application Serial No. 60/091,526, incorporated herein by reference.
- 5 The state of the art in DNA synthesis allows the artisan to construct *de novo* DNA compounds of size sufficient to construct a useful portion of a PKS module or domain. For purposes of the present invention, such synthetic DNA compounds are deemed to be a portion of a PKS.
- Thus, the hybrid modules of the invention are incorporated into a PKS to provide a hybrid PKS of the invention. A hybrid PKS of the invention can result not only:
- 10 (i) from fusions of heterologous domain (where heterologous means the domains in that module are from at least two different naturally occurring modules) coding sequences to produce a hybrid module coding sequence contained in a PKS gene whose product is incorporated into a PKS,
- 15 but also:

- (ii) from fusions of heterologous module (where heterologous module means two modules are adjacent to one another that are not adjacent to one another in naturally occurring PKS enzymes) coding sequences to produce a hybrid coding sequence contained in a PKS gene whose product is incorporated into a PKS,
- 20 (iii) from expression of one or more FK-520 PKS genes with one or more non-FK-520 PKS genes, including both naturally occurring and recombinant non-FK-520 PKS genes, and
- (iv) from combinations of the foregoing.

Various hybrid PKSs of the invention illustrating these various alternatives are described 25 herein.

Examples of the production of a hybrid PKS by co-expression of PKS genes from the FK-520 PKS and another non-FK-520 PKS include hybrid PKS enzymes produced by coexpression of FK-520 and rapamycin PKS genes. Preferably, such hybrid PKS enzymes are produced in recombinant *Streptomyces* host cells that produce FK-520 or FK-506 but 30 have been mutated to inactivate the gene whose function is to be replaced by the rapamycin PKS gene introduced to produce the hybrid PKS. Particular examples include (i) replacement of the *fkbC* gene with the *rapB* gene; and (ii) replacement of the *fkbA* gene with the *rapC* gene. The latter hybrid PKS produces 13,15-didesmethoxy-FK-520, if the host cell

is an FK-520 producing host cell, and 13,15-didesmethoxy-FK-506, if the host cell is an FK-506 producing host cell. The compounds produced by these hybrid PKS enzymes are immunosuppressants and neurotrophins but can be readily modified to act only as neurotrophins, as described in Example 6, below.

5 Other illustrative hybrid PKS enzymes of the invention are prepared by replacing the *fkbA* gene of an FK-520 or FK-506 producing host cell with a hybrid *fkbA* gene in which:

(a) the extender module 8 through 10, inclusive, coding sequences have been replaced by the coding sequences for extender modules 12 to 14, inclusive, of the rapamycin PKS; and

(b) the module 8 coding sequences have been replaced by the module 8 coding sequence of

10 the rifamycin PKS. When expressed with the other, naturally occurring FK-520 or FK-506 PKS genes and the genes of the modification enzymes, the resulting hybrid PKS enzymes produce, respectively, (a) 13-desmethoxy-FK-520 or 13-desmethoxy-FK-506; and (b) 13-desmethoxy-13-methyl-FK-520 or 13-desmethoxy-13-methyl-FK-506. In a preferred embodiment, these recombinant PKS genes of the invention are introduced into the

15 producing host cell by a vector such as pHU204, which is a plasmid pRMS derivative that has the well-characterized SCP2* replicon, the *colE1* replicon, the *tsr* and *bla* resistance genes, and a *cos* site. This vector can be used to introduce the recombinant *fkbA* replacement gene in an FK-520 or FK-506 producing host cell (or a host cell derived therefrom in which the endogenous *fkbA* gene has either been rendered inactive by

20 mutation, deletion or homologous recombination with the gene that replaces it) to produce the desired hybrid PKS.

In constructing hybrid PKSs of the invention, certain general methods may be helpful. For example, it is often beneficial to retain the framework of the module to be altered to make the hybrid PKS. Thus, if one desires to add DH and ER functionalities to a

25 module, it is often preferred to replace the KR domain of the original module with a KR, DH, and ER domain-containing segment from another module, instead of merely inserting DH and ER domains. One can alter the stereochemical specificity of a module by replacement of the KS domain with a KS domain from a module that specifies a different stereochemistry. See Lau *et al.*, 1999, "Dissecting the role of acyltransferase domains of

30 modular polyketide synthases in the choice and stereochemical fate of extender units," *Biochemistry* 38(5):1643-1651, incorporated herein by reference. Stereochemistry can also be changed by changing the KR domain. Also, one can alter the specificity of an AT domain by changing only a small segment of the domain. See Lau *et al.*, *supra*. One can

also take advantage of known linker regions in PKS proteins to link modules from two different PKSs to create a hybrid PKS. See Gokhale *et al.*, 16 Apr. 1999, "Dissecting and Exploiting Intermodular Communication in Polyketide Synthases," *Science* 284: 482-485, incorporated herein by reference.

5 The following Table lists references describing illustrative PKS genes and corresponding enzymes that can be utilized in the construction of the recombinant PKSs and the corresponding DNA compounds that encode them of the invention. Also presented are various references describing tailoring enzymes and corresponding genes that can be employed in accordance with the methods of the present invention.

10 **Avermectin**

U.S. Pat. No. 5,252,474 to Merck.

MacNeil *et al.*, 1993, Industrial Microorganisms: Basic and Applied Molecular Genetics, Baltz, Hegeman, & Skatrud, eds. (ASM), pp. 245-256, A Comparison of the Genes Encoding the Polyketide Synthases for Avermectin, Erythromycin, and Nemalectin.

15 MacNeil *et al.*, 1992, *Gene* 115: 119-125, Complex Organization of the *Streptomyces avermitilis* genes encoding the avermectin polyketide synthase.

Ikeda *et al.*, Aug. 1999, Organization of the biosynthetic gene cluster for the polyketide anthelmintic macrolide avermectin in *Streptomyces avermitilis*, *Proc. Natl. Acad. Sci. USA* 96: 9509-9514.

20 **Candididin (FR008)**

Hu *et al.*, 1994, *Mol. Microbiol.* 14: 163-172.

Epothilone

U.S. Pat. App. Serial No. 60/130,560, filed 22 April 1999.

Erythromycin

25 PCT Pub. No. 93/13663 to Abbott.

US Pat. No. 5,824,513 to Abbott.

Donadio *et al.*, 1991, *Science* 252:675-9.

Cortes *et al.*, 8 Nov. 1990, *Nature* 348:176-8, An unusually large multifunctional polypeptide in the erythromycin producing polyketide synthase of 30 *Saccharopolyspora erythraea*.

Glycosylation Enzymes

PCT Pat. App. Pub. No. 97/23630 to Abbott.

FK-506

Motamedi *et al.*, 1998, The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK-506, *Eur. J. biochem.* 256: 528-534.

Motamedi *et al.*, 1997, Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK-506, *Eur. J. Biochem.* 244: 74-80.

Methyltransferase

US 5,264,355, issued 23 Nov. 1993, Methylating enzyme from *Streptomyces* MA6858, 31-O-desmethyl-FK-506 methyltransferase.

Motamedi *et al.*, 1996, Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK-506 and FK-520, *J. Bacteriol.* 178: 5243-5248.

Streptomyces hygroscopicus

U.S. patent application Serial No. 09/154,083, filed 16 Sep. 1998.

Lovastatin

15 U.S. Pat. No. 5,744,350 to Merck.

Narbomycin

U.S. patent application Serial No. 60/107,093, filed 5 Nov. 1998, and Serial No. 60/120,254, filed 16 Feb. 1999.

Nemadectin

20 MacNeil *et al.*, 1993, *supra*.

Niddamycin

Kakavas *et al.*, 1997, Identification and characterization of the niddamycin polyketide synthase genes from *Streptomyces caelestis*, *J. Bacteriol.* 179: 7515-7522.

Oleandomycin

25 Swan *et al.*, 1994, Characterisation of a *Streptomyces antibioticus* gene encoding a type I polyketide synthase which has an unusual coding sequence, *Mol. Gen. Genet.* 242: 358-362.

U.S. patent application Serial No. 60/120,254, filed 16 Feb. 1999.

30 Olano *et al.*, 1998, Analysis of a *Streptomyces antibioticus* chromosomal region involved in oleandomycin biosynthesis, which encodes two glycosyltransferases responsible for glycosylation of the macrolactone ring, *Mol. Gen. Genet.* 259(3): 299-308.

Picromycin

PCT patent application US99/15047, filed 2 Jul. 1999.

Xue *et al.*, 1998, Hydroxylation of macrolactones YC-17 and narbomycin is mediated by the *pikC*-encoded cytochrome P450 in *Streptomyces venezuelae*. *Chemistry & Biology* 5(11): 661-667.

5 Xue *et al.*, Oct. 1998, A gene cluster for macrolide antibiotic biosynthesis in *Streptomyces venezuelae*: Architecture of metabolic diversity, *Proc. Natl. Acad. Sci. USA* 95: 12111-12116.

Platenolide

EP Pat. App. Pub. No. 791,656 to Lilly.

Rapamycin

10 Schwecke *et al.*, Aug. 1995, The biosynthetic gene cluster for the polyketide rapamycin, *Proc. Natl. Acad. Sci. USA* 92:7839-7843.

Aparicio *et al.*, 1996, Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase, *Gene* 169: 9-16.

15 **Rifamycin**

August *et al.*, 13 Feb. 1998, Biosynthesis of the ansamycin antibiotic rifamycin: deductions from the molecular analysis of the *rif* biosynthetic gene cluster of *Amycolatopsis mediterranei* S669, *Chemistry & Biology*. 5(2): 69-79.

Sorangium PKS

20 U.S. patent application Serial No. 09/144,085, filed 31 Aug. 1998.

Soraphen

U.S. Pat. No. 5,716,849 to Novartis.

Schupp *et al.*, 1995, *J. Bacteriology* 177: 3673-3679. A *Sorangium cellulosum* (Myxobacterium) Gene Cluster for the Biosynthesis of the Macrolide Antibiotic Soraphen

25 A: Cloning, Characterization, and Homology to Polyketide Synthase Genes from Actinomycetes.

Spiramycin

U.S. Pat. No. 5,098,837 to Lilly.

Activator Gene

30 U.S. Pat. No. 5,514,544 to Lilly.

Tylosin

EP Pub. No. 791,655 to Lilly.

U.S. Pat. No. 5,876,991 to Lilly.

Kuhstoss *et al.*, 1996, *Gene* 183:231-6.. Production of a novel polyketide through the construction of a hybrid polyketide synthase.

Tailoring enzymes

Merson-Davies and Cundliffe, 1994, *Mol. Microbiol.* 13: 349-355. Analysis of five 5 tylosin biosynthetic genes from the *tylBA* region of the *Streptomyces fradiae* genome.

As the above Table illustrates, there are a wide variety of polyketide synthase genes that serve as readily available sources of DNA and sequence information for use in constructing the hybrid PKS-encoding DNA compounds of the invention. Methods for 10 constructing hybrid PKS-encoding DNA compounds are described without reference to the FK-520 PKS in PCT patent publication No. 98/51695; U.S. Patent Nos. 5,672,491 and 5,712,146 and U.S. patent application Serial Nos. 09/073,538, filed 6 May 1998, and 09/141,908, filed 28 Aug 1998, each of which is incorporated herein by reference.

The hybrid PKS-encoding DNA compounds of the invention can be and often are 15 hybrids of more than two PKS genes. Moreover, there are often two or more modules in the hybrid PKS in which all or part of the module is derived from a second (or third) PKS. Thus, as one illustrative example, the present invention provides a hybrid FK-520 PKS that 20 contains the naturally occurring loading module and FkbP as well as modules one, two, four, six, seven, and eight, nine, and ten of the FK-520 PKS and further contains hybrid or heterologous modules three and five. Hybrid or heterologous module three contains an AT domain that is specific of methylmalonyl CoA and can be derived for example, from the erythromycin or rapamycin PKS genes. Hybrid or heterologous module five contains an AT 25 domain that is specific for malonyl CoA and can be derived for example, from the picromycin or rapamycin PKS genes.

While an important embodiment of the present invention relates to hybrid PKS 30 enzymes and corresponding genes, the present invention also provides recombinant FK-520 PKS genes in which there is no second PKS gene sequence present but which differ from the FK-520 PKS gene by one or more deletions. The deletions can encompass one or more modules and/or can be limited to a partial deletion within one or more modules. When a deletion encompasses an entire module, the resulting FK-520 derivative is at least two carbons shorter than the gene from which it was derived. When a deletion is within a module, the deletion typically encompasses a KR, DH, or ER domain, or both DH and ER domains, or both KR and DH domains, or all three KR, DH, and ER domains.

To construct a hybrid PKS or FK-520 derivative PKS gene of the invention, one can employ a technique, described in PCT Pub. No. 98/27203 and U.S. patent application Serial No. 08/989,332, filed 11 Dec. 1997, each of which is incorporated herein by reference, in which the large PKS gene is divided into two or more, typically three, segments, and each 5 segment is placed on a separate expression vector. In this manner, each of the segments of the gene can be altered, and various altered segments can be combined in a single host cell to provide a recombinant PKS gene of the invention. This technique makes more efficient the construction of large libraries of recombinant PKS genes, vectors for expressing those genes, and host cells comprising those vectors.

10 Thus, in one important embodiment, the recombinant DNA compounds of the invention are expression vectors. As used herein, the term expression vector refers to any nucleic acid that can be introduced into a host cell or cell-free transcription and translation medium. An expression vector can be maintained stably or transiently in a cell, whether as part of the chromosomal or other DNA in the cell or in any cellular compartment, such as a 15 replicating vector in the cytoplasm. An expression vector also comprises a gene that serves to produce RNA that is translated into a polypeptide in the cell or cell extract. Furthermore, expression vectors typically contain additional functional elements, such as resistance-conferring genes to act as selectable markers.

20 The various components of an expression vector can vary widely, depending on the intended use of the vector. In particular, the components depend on the host cell(s) in which the vector will be used or is intended to function. Vector components for expression and maintenance of vectors in *E. coli* are widely known and commercially available, as are vector components for other commonly used organisms, such as yeast cells and 25 *Streptomyces* cells.

25 In a preferred embodiment, the expression vectors of the invention are used to construct recombinant *Streptomyces* host cells that express a recombinant PKS of the invention. Preferred *Streptomyces* host cell/vector combinations of the invention include *S. coelicolor* CH999 and *S. lividans* K4-114 host cells, which do not produce actinorhodin, and expression vectors derived from the pRM1 and pRM5 vectors, as described in U.S. 30 Patent No. 5,830,750 and U.S. patent application Serial Nos. 08/828,898, filed 31 Mar. 1997, and 09/181,833, filed 28 Oct. 1998, each of which is incorporated herein by reference.

The present invention provides a wide variety of expression vectors for use in *Streptomyces*. For replicating vectors, the origin of replication can be, for example and without limitation, a low copy number vector, such as SCP2* (see Hopwood *et al.*, *Genetic Manipulation of Streptomyces: A Laboratory manual* (The John Innes Foundation, Norwich, U.K., 1985); Lydiate *et al.*, 1985, *Gene* 35: 223-235; and Kieser and Melton, 1988, *Gene* 65: 83-91, each of which is incorporated herein by reference), SLP1.2 (Thompson *et al.*, 1982, *Gene* 20: 51-62, incorporated herein by reference), and SG5(ts) (Muth *et al.*, 1989, *Mol. Gen. Genet.* 219: 341-348, and Bierman *et al.*, 1992, *Gene* 116: 43-49, each of which is incorporated herein by reference), or a high copy number vector, such as pIJ101 and pJV1 (see Katz *et al.*, 1983, *J. Gen. Microbiol.* 129: 2703-2714; Vara *et al.*, 1989, *J. Bacteriol.* 171: 5782-5781; and Servin-Gonzalez, 1993, *Plasmid* 30: 131-140, each of which is incorporated herein by reference). Generally, however, high copy number vectors are not preferred for expression of genes contained on large segments of DNA. For non-replicating and integrating vectors, it is useful to include at least an *E. coli* origin of replication, such as from pUC, p1P, p1I, and pBR. For phage based vectors, the phages phiC31 and KC515 can be employed (see Hopwood *et al.*, *supra*).

Typically, the expression vector will comprise one or more marker genes by which host cells containing the vector can be identified and/or selected. Useful antibiotic resistance conferring genes for use in *Streptomyces* host cells include the *ermE* (confers resistance to erythromycin and other macrolides and lincomycin), *tsr* (confers resistance to thiostrepton), *aadA* (confers resistance to spectinomycin and streptomycin), *aacC4* (confers resistance to apramycin, kanamycin, gentamicin, geneticin (G418), and neomycin), *hyg* (confers resistance to hygromycin), and *vph* (confers resistance to viomycin) resistance conferring genes.

The recombinant PKS gene on the vector will be under the control of a promoter, typically with an attendant ribosome binding site sequence. The present invention provides the endogenous promoters of the FK-520 PKS and related biosynthetic genes in recombinant form, and these promoters are preferred for use in the native hosts and in heterologous hosts in which the promoters function. A preferred promoter of the invention is the *fkbO* gene promoter, comprised in a sequence of about 270 bp between the start of the open reading frames of the *fkbO* and *fkbB* genes. The *fkbO* promoter is believed to be bi-directional in that it promotes transcription of the genes *fkbO*, *fkbP*, and *fkbA* in one direction and *fkbB*, *fkbC*, and *fkbL* in the other. Thus, in one aspect, the present invention

provides a recombinant expression vector comprising the promoter of the *fkbO* gene of an FK-520 producing organism positioned to transcribe a gene other than *fkbO*. In a preferred embodiment the transcribed gene is an FK-520 PKS gene. In another preferred embodiment, the transcribed gene is a gene that encodes a protein comprised in a hybrid PKS.

5 Heterologous promoters can also be employed and are preferred for use in host cells in which the endogenous FK-520 PKS gene promoters do not function or function poorly. A preferred heterologous promoter is the *actI* promoter and its attendant activator gene *actII-ORF4*, which is provided in the pRM1 and pRM5 expression vectors, *supra*. This promoter is activated in the stationary phase of growth when secondary metabolites are normally 10 synthesized. Other useful *Streptomyces* promoters include without limitation those from the *ermE* gene and the *melC1* gene, which act constitutively, and the *tipA* gene and the *merA* gene, which can be induced at any growth stage. In addition, the T7 RNA polymerase system has been transferred to *Streptomyces* and can be employed in the vectors and host 15 cells of the invention. In this system, the coding sequence for the T7 RNA polymerase is inserted into a neutral site of the chromosome or in a vector under the control of the inducible *merA* promoter, and the gene of interest is placed under the control of the T7 promoter. As noted above, one or more activator genes can also be employed to enhance the activity of a promoter. Activator genes in addition to the *actII-ORF4* gene discussed above 20 include *dnrI*, *redD*, and *ptpA* genes (see U.S. patent application Serial No. 09/181,833, *supra*) to activate promoters under their control.

25 In addition to providing recombinant DNA compounds that encode the FK-520 PKS, the present invention also provides DNA compounds that encode the ethylmalonyl CoA and 2-hydroxymalonyl CoA utilized in the synthesis of FK-520. Thus, the present invention also provides recombinant host cells that express the genes required for the biosynthesis of ethylmalonyl CoA and 2-hydroxymalonyl CoA. Figures 3 and 4 show the 30 location of these genes on the cosmids of the invention and the biosynthetic pathway that produces ethylmalonyl CoA.

For 2-hydroxymalonyl CoA biosynthesis, the *fkbH*, *fkbI*, *fkbJ*, and *fkbK* genes are sufficient to confer this ability on *Streptomyces* host cells. For conversion of 2-hydroxymalonyl to 2-methoxymalonyl, the *fkbG* gene is also employed. While the complete coding sequence for *fkbH* is provided on the cosmids of the invention, the sequence for this gene provided herein may be missing a T residue, based on a comparison made with a similar gene cloned from the ansamitocin gene cluster by Dr. H. Floss. Where the sequence

herein shows one T, there may be two, resulting in an extension of the *fkbH* reading frame to encode the amino acid sequence:

MTIVKCLVWDLNDNTLWRGTVLEDDEVVLTDEIREVITLDDRGILQAVASKNDHD
LAWERLERLGVAEYFVLARIGWGPKSQSVREIATELNFAPTTIAFIDDQPAERAEV
5 FHLPEVRCYPAEQQAATLLSLPEFSPPVSTVDSRRRLMYQAGFARDQAREAYSGPD
EDFLRSLDLSMTIAPAGEEELSVEELRLTSQMNATGVHYSADLRALLTDAHE
VLVVTMGDRFGPHGAVGIILLEKKPSTWHLKLLATSCRVVSFGAGATILNWLTDQG
ARAGAHLVADFRRTDRNRMMEIAYRFAGFADSDCPCVSEVAGASAAGVERLHLEP
SARPAPTTLTAADIAPVTVSAAG.

10 For ethylmalonyl CoA biosynthesis, one requires only a crotonyl CoA reductase, which can be supplied by the host cell but can also be supplied by recombinant expression of the *fkbs* gene of the present invention. To increase yield of ethylmalonyl CoA, one can also express the *fkbe* and *fkbu* genes as well. While such production can be achieved using only the recombinant genes above, one can also achieve such production by placing into the
15 recombinant host cell a large segment of the DNA provided by the cosmids of the invention. Thus, for 2-hydroxymalonyl and 2-methoxymalonyl CoA biosynthesis, one can simply provide the cells with the segment of DNA located on the left side of the FK-520 PKS genes shown in Figure 1. For ethylmalonyl CoA biosynthesis, one can simply provide the cells with the segment of DNA located on the right side of the FK-520 PKS genes shown in
20 Figure 1 or, alternatively, both the right and left segments of DNA.

The recombinant DNA expression vectors that encode these genes can be used to construct recombinant host cells that can make these important polyketide building blocks from cells that otherwise are unable to produce them. For example, *Streptomyces coelicolor* and *Streptomyces lividans* do not synthesize ethylmalonyl CoA or 2-hydroxymalonyl CoA.
25 The invention provides methods and vectors for constructing recombinant *Streptomyces coelicolor* and *Streptomyces lividans* that are able to synthesize either or both ethylmalonyl CoA and 2-hydroxymalonyl CoA. These host cells are thus able to make polyketides, those requiring these substrates, that cannot otherwise be made in such cells.

30 In a preferred embodiment, the present invention provides recombinant *Streptomyces* host cells, such as *S. coelicolor* and *S. lividans*, that have been transformed with a recombinant vector of the invention that codes for the expression of the ethylmalonyl CoA biosynthetic genes. The resulting host cells produce ethylmalonyl CoA and so are preferred host cells for the production of polyketides produced by PKS enzymes that

comprise one or more AT domains specific for ethylmalonyl CoA. Illustrative PKS enzymes of this type include the FK-520 PKS and a recombinant PKS in which one or more AT domains is specific for ethylmalonyl CoA.

In a related embodiment, the present invention provides *Streptomyces* host cells in which one or more of the ethylmalonyl or 2-hydroxymalonyl biosynthetic genes have been deleted by homologous recombination or rendered inactive by mutation. For example, deletion or inactivation of the *fkbG* gene can prevent formation of the methoxyl groups at C-13 and C-15 of FK-520 (or, in the corresponding FK-506 producing cell, FK-506), leading to the production of 13,15-didesmethoxy-13,15-dihydroxy-FK-520 (or, in the corresponding FK-506 producing cell, 13,15-didesmethoxy-13,15-dihydroxy-FK-506). If the *fkbG* gene product acts on 2-hydroxymalonyl and the resulting 2-methoxymalonyl substrate is required for incorporation by the PKS, the AT domains of modules 7 and 8 may bind malonyl CoA and methylmalonyl CoA. Such incorporation results in the production of a mixture of polyketides in which the methoxy groups at C-13 and C-15 of FK-520 (or FK-506) are replaced by either hydrogen or methyl.

This possibility of non-specific binding results from the construction of a hybrid PKS of the invention in which the AT domain of module 8 of the FK-520 PKS replaced the AT domain of module 6 of DEBS. The resulting PKS produced, in *Streptomyces lividans*, 6-dEB and 2-desmethyl-6-dEB, indicating that the AT domain of module 8 of the FK-520 PKS could bind malonyl CoA and methylmalonyl CoA substrates. Thus, one could possibly also prepare the 13,15-didesmethoxy-FK-520 and corresponding FK-506 compounds of the invention by deleting or otherwise inactivating one or more or all of the genes required for 2-hydroxymalonyl CoA biosynthesis, i.e., the *fkbH*, *fkbI*, *fkbJ*, and *fkbK* genes. In any event, the deletion or inactivation of one or more biosynthetic genes required for ethylmalonyl and/or 2-hydroxymalonyl production prevents the formation of polyketides requiring ethylmalonyl and/or 2-hydroxymalonyl for biosynthesis, and the resulting host cells are thus preferred for production of polyketides that do not require the same.

The host cells of the invention can be grown and fermented under conditions known in the art for other purposes to produce the compounds of the invention. See, e.g., U.S. Patent Nos. 5,194,378; 5,116,756; and 5,494,820, incorporated herein by reference, for suitable fermentation processes. The compounds of the invention can be isolated from the fermentation broths of these cultured cells and purified by standard procedures. Preferred compounds of the invention include the following compounds: 13-desmethoxy-FK-506; 13-

desmethoxy-FK-520; 13,15-didesmethoxy-FK-506; 13,15-didesmethoxy-FK-520; 13-desmethoxy-18-hydroxy-FK-506; 13-desmethoxy-18-hydroxy-FK-520; 13,15-didesmethoxy-18-hydroxy-FK-506; and 13,15-didesmethoxy-18-hydroxy-FK-520. These compounds can be further modified as described for tacrolimus and FK-520 in U.S. Patent 5 Nos. 5,225,403; 5,189,042; 5,164,495; 5,068,323; 4,980,466; and 4,920,218, incorporated herein by reference.

Other compounds of the invention are shown in Figure 8, Parts A and B. In Figure 8, Part A, illustrative C-32-substituted compounds of the invention are shown in two columns under the heading R. The substituted compounds are preferred for topical administration 10 and are applied to the dermis for treatment of conditions such as psoriasis. In Figure 8, Part B, illustrative reaction schemes for making the compounds shown in Figure 8, Part A, are provided. In the upper scheme in Figure 8, Part B, the C-32 substitution is a tetrazole moiety, illustrative of the groups shown in the left column under R in Figure 8, Part A. In the lower scheme in Figure 8, Part B, the C-32 substitution is a disubstituted amino group, 15 where R₃ and R₄ can be any group similar to the illustrative groups shown attached to the amine in the right column under R in Figure 8, Part A. While Figure 8 shows the C-32-substituted compounds in which the C-15-methoxy is present, the invention includes these C-32-substituted compounds in which C-15 is ethyl, methyl, or hydrogen. Also, while C-21 is shown as substituted with ethyl or allyl, the compounds of the invention includes the C- 20 32-substituted compounds in which C-21 is substituted with hydrogen or methyl.

To make these C-32-substituted compounds, Figure 8, Part B, provides illustrative reaction schemes. Thus, a selective reaction of the starting compound (see Figure 8, Part B, for an illustrative starting compound) with trifluoromethanesulfonic anhydride in the presence of a base yields the C-32 O-triflate derivative, as shown in the upper scheme of 25 Figure 8, Part B. Displacement of the triflate with 1H-tetrazole or triazole derivatives provides the C-32 tetrazole or triazole derivative. As shown in the lower scheme of Figure 8, Part B, reacting the starting compound with p-nitrophenylchloroformate yields the corresponding carbonate, which, upon displacement with an amino compound, provides the corresponding carbamate derivative.

30 The compounds can be readily formulated to provide the pharmaceutical compositions of the invention. The pharmaceutical compositions of the invention can be used in the form of a pharmaceutical preparation, for example, in solid, semisolid, or liquid form. This preparation contains one or more of the compounds of the invention as an active

ingredient in admixture with an organic or inorganic carrier or excipient suitable for external, enteral, or parenteral application. The active ingredient may be compounded, for example, with the usual non-toxic, pharmaceutically acceptable carriers for tablets, pellets, capsules, suppositories, solutions, emulsions, suspensions, and any other form suitable for use. Suitable formulation processes and compositions for the compounds of the present invention are described with respect to tacrolimus in U.S. Patent Nos. 5,939,427; 5,922,729; 5,385,907; 5,338,684; and 5,260,301, incorporated herein by reference. Many of the compounds of the invention contain one or more chiral centers, and all of the stereoisomers are included within the scope of the invention, as pure compounds as well as mixtures of stereoisomers. Thus the compounds of the invention may be supplied as a mixture of stereoisomers in any proportion.

The carriers which can be used include water, glucose, lactose, gum acacia, gelatin, mannitol, starch paste, magnesium trisilicate, talc, corn starch, keratin, colloidal silica, potato starch, urea, and other carriers suitable for use in manufacturing preparations, in solid, semi-solid, or liquified form. In addition, auxiliary stabilizing, thickening, and coloring agents and perfumes may be used. For example, the compounds of the invention may be utilized with hydroxypropyl methylcellulose essentially as described in U.S. Patent No. 4,916,138, incorporated herein by reference, or with a surfactant essentially as described in EPO patent publication No. 428,169, incorporated herein by reference.

Oral dosage forms may be prepared essentially as described by Hondo *et al.*, 1987, *Transplantation Proceedings XIX*, Supp. 6: 17-22, incorporated herein by reference. Dosage forms for external application may be prepared essentially as described in EPO patent publication No. 423,714, incorporated herein by reference. The active compound is included in the pharmaceutical composition in an amount sufficient to produce the desired effect upon the disease process or condition.

For the treatment of conditions and diseases relating to immunosuppression or neuronal damage, a compound of the invention may be administered orally, topically, parenterally, by inhalation spray, or rectally in dosage unit formulations containing conventional non-toxic pharmaceutically acceptable carriers, adjuvant, and vehicles. The term parenteral, as used herein, includes subcutaneous injections, and intravenous, intramuscular, and intrasternal injection or infusion techniques.

Dosage levels of the compounds of the present invention are of the order from about 0.01 mg to about 50 mg per kilogram of body weight per day, preferably from about 0.1 mg

to about 10 mg per kilogram of body weight per day. The dosage levels are useful in the treatment of the above-indicated conditions (from about 0.7 mg to about 3.5 mg per patient per day, assuming a 70 kg patient). In addition, the compounds of the present invention may be administered on an intermittent basis, i.e., at semi-weekly, weekly, semi-monthly, or monthly intervals.

The amount of active ingredient that may be combined with the carrier materials to produce a single dosage form will vary depending upon the host treated and the particular mode of administration. For example, a formulation intended for oral administration to humans may contain from 0.5 mg to 5 g of active agent compounded with an appropriate and convenient amount of carrier material, which may vary from about 5 percent to about 95 percent of the total composition. Dosage unit forms will generally contain from about 0.5 mg to about 500 mg of active ingredient. For external administration, the compounds of the invention can be formulated within the range of, for example, 0.00001% to 60% by weight, preferably from 0.001% to 10% by weight, and most preferably from about 0.005% to 0.8% by weight. The compounds and compositions of the invention are useful in treating disease conditions using doses and administration schedules as described for tacrolimus in U.S. Patent Nos. 5,542,436; 5,365,948; 5,348,966; and 5,196,437, incorporated herein by reference. The compounds of the invention can be used as single therapeutic agents or in combination with other therapeutic agents. Drugs that can be usefully combined with compounds of the invention include one or more immunosuppressant agents such as rapamycin, cyclosporin A, FK-506, or one or more neurotrophic agents.

It will be understood, however, that the specific dosage level for any particular patient will depend on a variety of factors. These factors include the activity of the specific compound employed; the age, body weight, general health, sex, and diet of the subject; the time and route of administration and the rate of excretion of the drug; whether a drug combination is employed in the treatment; and the severity of the particular disease or condition for which therapy is sought.

A detailed description of the invention having been provided above, the following examples are given for the purpose of illustrating the present invention and shall not be construed as being a limitation on the scope of the invention or claims.

Example 1

Replacement of Methoxyl with Hydrogen or Methyl at C-13 of FK-520

The C-13 methoxyl group is introduced into FK-520 via an AT domain in extender module 8 of the PKS that is specific for hydroxymalonyl and by methylation of the hydroxyl group by an S-adenosyl methionine (SAM) dependent methyltransferase. Metabolism of FK-506 and FK-520 primarily involves oxidation at the C-13 position into 5 an inactive derivative that is further degraded by host P450 and other enzymes. The present invention provides compounds related in structure to FK-506 and FK-520 that do not contain the C-13 methoxy group and exhibit greater stability and a longer half-life *in vivo*. These compounds are useful medicaments due to their immunosuppressive and 10 neurotrophic activities, and the invention provides the compounds in purified form and as pharmaceutical compositions.

The present invention also provides the novel PKS enzymes that produce these novel compounds as well as the expression vectors and host cells that produce the novel PKS enzymes. The novel PKS enzymes include, among others, those that contain an AT domain specific for either malonyl CoA or methylmalonyl CoA in module 8 of the FK-506 15 and FK-520 PKS. This example describes the construction of recombinant DNA compounds that encode the novel FK-520 PKS enzymes and the transformation of host cells with those recombinant DNA compounds to produce the novel PKS enzymes and the polyketides produced thereby.

To construct an expression cassette for performing module 8 AT domain 20 replacements in the FK-520 PKS, a 4.6 kb *Sph*I fragment from the FK-520 gene cluster was cloned into plasmid pLitmus 38 (a cloning vector available from New England Biolabs). The 4.6 kb *Sph*I fragment, which encodes the ACP domain of module 7 followed by module 8 through the KR domain, was isolated from an agarose gel after digesting the cosmid 25 pKOS65-C31 with *Sph* I. The clone having the insert oriented so the single *Sac*I site was nearest to the *Spe*I end of the polylinker was identified and designated as plasmid pKOS60- 21-67. To generate appropriate cloning sites, two linkers were ligated sequentially as 30 follows. First, a linker was ligated between the *Spe*I and *Sac*I sites to introduce a *Bgl*II site at the 5' end of the cassette, to eliminate interfering polylinker sites, and to reduce the total insert size to 4.5 kb (the limit of the phage KC515). The ligation reactions contained 5 picomolar unphosphorylated linker DNA and 0.1 picomolar vector DNA, i.e., a 50-fold molar excess of linker to vector. The linker had the following sequence:

5'-CTAGTGGCAGATCTGGCAGCT-3'
3'-ACCCGTCTAGACCG-5'

The resulting plasmid was designated pKOS60-27-1.

Next, a linker of the following sequence was ligated between the unique *Sph*I and *Af*II sites of plasmid pKOS60-27-1 to introduce an *Nsi*I site at the 3' end of the module 8 cassette. The linker employed was:

5 5'-GGGATGCATGGC-3'
 3'-GTACCCCTACGTACCGAATT-5'

The resulting plasmid was designated pKOS60-29-55.

To allow in-frame insertions of alternative AT domains, sites were engineered at the 5' end (*Avr* II or *Nhe* I) and 3' end (*Xho* I) of the AT domain using the polymerase chain reaction (PCR) as follows. Plasmid pKOS60-29-55 was used as a template for the PCR and 10 sequence 5' to the AT domain was amplified with the primers *SpeBgl*-fwd and either *Avr*-rev or *Nhe*-rev:

SpeBgl-fwd 5'-CGACTCACTAGTGGGCAGATCTGG-3'
Avr-rev 5'-CACGCCTAGGCCGGTCGGTCTCGGGCCAC-3'
Nhe-rev 5'-GCGGCTAGCTGCTCGCCCATCGCGGGATGC-3'

15 The PCR included, in a 50 μ l reaction, 5 μ l of 10x *Pfu* polymerase buffer (Stratagene), 5 μ l 10x z-dNTP mixture (2 mM dATP, 2 mM dCTP, 2 mM dTTP, 1 mM dGTP, 1 mM 7-deaza-GTP), 5 μ l DMSO, 2 μ l of each primer (10 μ M), 1 μ l of template DNA (0.1 μ g/ μ l), and 1 μ l of cloned *Pfu* polymerase (Stratagene). The PCR conditions were 95°C for 2 min., 25 cycles at 95°C for 30 sec., 60°C for 30 sec., and 72°C for 4 min., 20 followed by 4 min. at 72°C and a hold at 0°C. The amplified DNA products and the Litmus vectors were cut with the appropriate restriction enzymes (*Bgl*II and *Avr*II or *Spe*I and *Nhe*I), and cloned into either pLitmus 28 or pLitmus38 (New England Biolabs), respectively, to generate the constructs designated pKOS60-37-4 and pKOS60-37-2, respectively.

25 Plasmid pKOS60-29-55 was again used as a template for PCR to amplify sequence 3' to the AT domain using the primers *BsrXho*-fwd and *NsiAfl*-rev:

BsrXho-fwd 5'-GATGTACAGCTCGAGTCGGCACGCCGGCGCATC-3'
NsiAfl-rev 5'-CGACTCACTTAAGCCATGCATCC-3'

30 PCR conditions were as described above. The PCR fragment was cut with *Bsr*GI and *Af*II, gel isolated, and ligated into pKOS60-37-4 cut with *Asp*718 and *Af*II and inserted into pKOS60-37-2 cut with *Bsr*GI and *Af*II, to give the plasmids pKOS60-39-1 and pKOS60-39-13, respectively. These two plasmids can be digested with *Avr*II and *Xhol* or *Nhe*I and *Xhol*, respectively, to insert heterologous AT domains specific for malonyl, methylmalonyl, ethylmalonyl, or other extender units.

Malonyl and methylmalonyl-specific AT domains were cloned from the rapamycin cluster using PCR amplification with a pair of primers that introduce an *AvrII* or *NheI* site at the 5' end and an *XbaI* site at the 3' end. The PCR conditions were as given above and the primer sequences were as follows:

5

RATN1 5'-ATCCTAGGCAGGCRGGYGTGTCGTCCTTCGG-3'
(3' end of Rap KS sequence and universal for malonyl and methylmalonyl CoA).

RATMN2 5'-ATGCTAGCCGCCGTTCCCCGTCTCGCGCG-3'

(Rap AT shorter version 5'- sequence and specific for malonyl CoA),

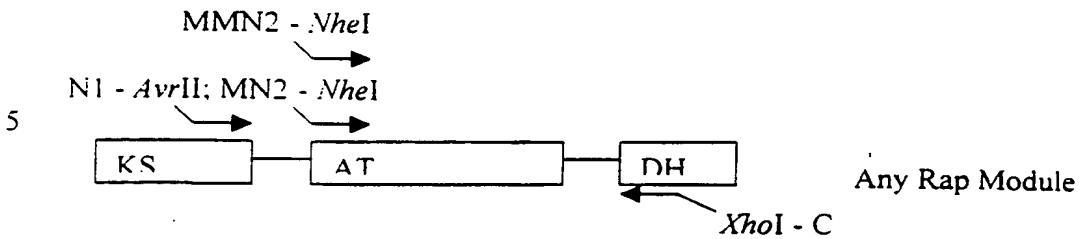
10

RATMMN2 5'-ATGCTAGCGGATTCTCGTCGGTGGTTCGCCGA-3'

(Rap AT shorter version 5'- sequence and specific for methylmalonyl CoA), and

RATC 5'-ATCTCGAGCCAGTASCGCTGGTGYTGGAAAGG-3'

(Rap DH 5'- sequence and universal for malonyl and methylmalonyl CoA).



10 Because of the high sequence similarity in each module of the rapamycin cluster, each primer was expected to prime any of the AT domains. PCR products representing ATs specific for malonyl or methylmalonyl extenders were identified by sequencing individual cloned PCR products. Sequencing also confirmed that the chosen clones contained no cloning artifacts. Examples of hybrid modules with the rapamycin AT12 and AT13 domains are shown in a separate figure.

15 The *AvrII-Xhol* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 12 of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below. The AT of rap module 12 is specific for incorporation of malonyl units.

20 AGATCTGGCAGCTGCCGAAGCGCTGCTGACGCTCGTCCGGAGAGCACC 50
 I W Q L A E A L L T L V R E S T
 GCCGCCGTGCTCGGCCACGTGGTGGCGAGGACATCCCCGCGACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTACCCGCGTCCAGCTGCGCAACG 150
 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGCACCCGGTGTGCGGCTGAACGCCACGGCGGTCTTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTCGCCGGGAAGCTCGGCGACGAACGTACCGG 250
 F P T P H V L A G K L G D E L T G
 CACCCGCGCGCCCGTCGTGCCCGGACCGCGGCCACGGCCGGTGCACG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCCTGCCGGCTGCCCGGCGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGGCCAT 400
 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCCGGCTGGACGGTGCACGGATCTACGACC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCCGACCGCATCGGCAAGACCTTCGTCGGCACGGTGGCTTCCTC 500
 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGCTTCGACGCCGGCTCTCGGCATCAGCCCGCGCA 550
 T G A T T G F D A A F F G I S P R E
 GGCCCTCGCGATGGACCCCGCAGCAGCGGTGCTCCTGGAGACGTCGTGGG 600
 A L A M D P Q Q R V L L E T S W
 AGGCCTTCGAAAGCGCCGGCATACCCCGGACTCGACCCCGGGCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCCTTCCTACGGTTACGGCACCGGTGGCA 700
 T G V F Y S A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGCGACCCGGCTCGCAGACCAAGTGTGCTCTCGGGCC 750
 T D G F G A T G S Q T S V L S G
 GGCTGTCGTACTTCTACGGTCTGGAGGGTCCGGCGGTACGGTCGACACG 800

R L S Y F Y G L E G P A V T V D T
 SCGTGTTCGTCGCTGGTGGCGCTGCACCAAGGCCGGCAGTCCCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCCGCGAATGCTCGCTCGCCCTGGTCCGGCGGCTCACGGTGAATGGT 900
 5 S G E C S L A L V G G V T V M A
 CTCCCGCGCGCTTCTGGAGTTCTCCCGCAGCGCGCCCTCGCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 GGCCCGGGCGAAGCGGTTCCGGCGGGTGCGGACGGCACGAGCTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 10 GGGTGCCGCTGTGCTGATCGTCGAGAGGCTCTCCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTCACACCGTCCCTGGCGGTGTCGCGGTCAACCAGGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 CCCTCCAACGGGCTGCGCGCCGAACGGGCCGTCGAGGAGCGGGTGT 1150
 15 A S N G L S A P N G P S Q E R V I
 CCGGCAGGCCCTGGCCAACGCCGGGCTCACCCCGGCGGACGTGGACGCC 1200
 R Q A L A N A G L T P A D V D A
 TCGAGGCCAACGGCACCGGACCCAGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 20 GCGGTACTGGCACCTACGGACAGGAGCGCGCCACCCCCCTGCTGCTGG 1300
 A V L A T Y G Q E R A T P L I L G
 CTCGCTGAAGTCAAACATCGGCCACGCCAACGGCCGTCGGCGCTCGCC 1350
 S L K S N I G H A Q A A S G V A
 SCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGGAGCTGCCGCCGACG 1400
 25 G I I K M V Q A L R H G E L P P T
 CTGCACGCCACGAGCCGTCGCCACGTGCACTGGACGCCGGCGCGT 1450
 L H A D E P S P H V D W T A G A V
 CGAACTGCTGACGTGCCCGGCGTGGCCCGAGACCGACCCGCTAGGC 1500
 E L L T S A R P W P E T D R P R
 30 GGGCAGGCGTGTGCTCTCGGGATCAGTGGCACCAACGCCACGTCACTC 1550
 R A G V S S F G I S G T N A H V I
 CTGGAAAGCCACCCCCCACTCAGCCTGCGACAACCGGGTGTGAGCG 1600
 L E S A P P T Q P A D N A V I E R
 GGCACCGGAGTGGTGGCGTTGGTGAATTCGGCCAGGACCCAGTCGGCTT 1650
 35 A P E W V P L V I S A R T Q S A
 TGACTGAGCACCGAGGGCCGGTGGCTGCGTATCTGGCGCGTCGCCCGGG 1700
 L T E H E G R L R A Y L A A S P G
 GTGGATATGCGGGCTGCGCATCGACGCTGGCGATGACACGGTCGGT 1750
 V D M R A V A S T L A M T R S V F
 40 CGAGCACCGTGCCGTGCTGGAGATGACACCGTCACCGCACCGCTG 1800
 E H P A V L L G D D T V T G T A
 TGTCTGACCCCTCGGGCGGTGTTGCTTCCCGGACAGGGGTCGAGCGT 1850
 V S D P R A V F V F P G Q G S Q R
 45 GCTGGCATGGTGAGGAACCTGGCCGCCGTTCCCGTCTCGCGCGGAT 1900
 A G M G E E L A A A F P V F A R I
 CCATCAGCAGGTGTGGACCTGCTGATGTGCCGATCTGGAGGTGAACG 1950
 H Q Q V W D L L D V P D L E V N
 AGACCCGGTTACGCCAGCCGGCTGTCGAATGCAGGTGGCTCTGTT 2000
 E T G Y A Q P A L F A M Q V A L F
 50 GGGCTGCTGGAATCGTGGGTGTACGACCGGACGCCGGTGTGATGCCATT 2050
 G L L E S W G V R P D A V I G H S
 GCTGGGTGAGCTTGCCTGCGTATGTGTCCGGGGTGTGGTCGTTGGAGG 2100
 V G E L A A A Y V S G V W S L E
 ATGCCCTGCACCTTGGTGTCCGGCGCCGCTCGTCTGATGCAGGCTCTGCC 2150
 55 S A C T L V S A R A R L M Q A L P
 CGGGGTGGGGTGTGGCTGCTGCCGGTCTCGGAGGTGAGGCCGGGC 2200
 A G G V X V A V P V S E D E A R A
 CGTGGCTGGGTGAGGGTGTGGAGATGCCGGCGTCAACGGCCCGTGGT 2250
 V L G E G V E I A A V N G P S S
 TCGTTCTCTCCGGTGTGATGAGGCCGGCGTGTGAGGCCGGAGGGGCTG 2300

I V L S C D E A A V I Q A A E G L
 GGGAAAGTGGACGGGCTGGCGACCAGCCACGCCGTTCCATTCCGCCCTAT 2350
 G K W T R L A T S H A F H S A F M
 GGAACCCATGGCTGGAGGGAGTTCCGGCGGTGCGCGAAGGCCTGACCTACG 2400
 5 E P M L E E F P A V A E G L T Y
 GGACGCCGCAGGTCTCCATGGCCGTTGGTGTACAGGTGACCACCGCTGAG 2450
 R T P Q V S M A V G D Q V T T A E
 TACTGGGTGCGGCAGGTCCGGGACACGGTCGGTCCGGCGAGCAGGTGGC 2500
 Y W V R Q V R D T V R F G E Q V A
 10 CTCGTACGAGGACGCCGTGTCGAGCTGGGTGCCGACCGGTCACTGG 2550
 S Y E D A V F V E L G A D R S L
 CCCGCCTGGTCGACGGTGTGCGATGCTGCACGGCGACCACGAAATCCAG 2600
 A R L V D G V A M L H G D H E I Q
 15 GCGCGATCGGCGCCCTGGCCACCTGTATGTCAACGGCGTACGGTCGA 2650
 A A I G A L A H L Y V N G V T V D
 CTGGCCCGCGCTCCTGGCGATGCTCCGGCAACACGGGTGCTGGACCTTC 2700
 W P A L L G D A P A T R V L D L
 CGACATACGCCCTCCAGCACCGCCTACTGGCTCGAGTCGGCACGCCCG 2750
 P T Y A F Q H Q R Y W L E S A R P
 20 GCCGCATCCGACGCCGGCCACCCGTGCTGGGCTCCGGTATGCCCTCGC 2800
 A A S D A G H P V L G S G I A L A
 CGGGTCGCCGGGCCGGTGTTCACGGGTTCCGTGCCGACCGGTGCGGAC 2850
 G S P G R V F T G S V P T G A D
 25 GCGCGGTGTTCGTCGCGAGCTGGCGCTGGCGCCGGACCGCGTGCAC 2900
 R A V F V A E L A L A A A D A V D
 TCGGCCACGGTCGAGCGGCTCGACATGCCCTCGTGCCGGCGCCGGGG 2950
 C A T V E R L D I A S V P G R P G
 CCATGGCCGACGACCGTACAGACCTGGGTGACGAGCCGGACGACG 3000
 H G R T T V Q T W V D E P A D D
 30 GCCGGCGCCGGTTACCGTGACACCCGCACCGCGACGCCCGTGGACG 3050
 G R R R F T V H T R T G D A P W T
 CTGCACGCCAGGGGGGTGCTGCGCCCCCATGGCACGGCCCTGCCGATGC 3100
 L H A E G V L R P H G T A L P D A
 GGGCGACGCCGAGTGGCCCCCACCGGGCGCGGTGCCCGCGGACGGCTGC 3150
 35 A D A E W P P P G A V P A D G L
 CGGGTGTGGCGCCGGGGGACAGGTCTCGCCGAGGCCGAGGTGGAC 3200
 P G V W R R G D Q V F A E A E V D
 GGACCGGACGGTTTCGTGGTGCACCCGACCTGCTCGACGCCGTCTTC 3250
 G P D G F V V H P D L L D A V F S
 40 CGGGTGGCGACGGAAAGCCGCCAGCCGGATGSGCGGACCTGACGG 3300
 A V G D G S R Q P A G W R D L T
 TGCACCGCTGGACGCCACCGTACTGGCGCCCTGCCCTACCCGGCCACC 3350
 V H A S D A T V L R A C L T R R T
 GACGGAGCCATGGGATTGCGCCCTCGACGGCGCCGCTGCCGGTACT 3400
 45 D G A M G F A A F D G A G L P V L
 CACCGCGGAGGCGGTGACGCTGGGGAGGTGGCGTCACCGTCCGGCTCCG 3450
 T A E A V T L R E V A S P S G S
 AGGAGTCGGACGGCTGCACCGGTTGGAGTGGCTGCCGGTCGCCGAGGCG 3500
 E E S D G L H R L E W L A V A E A
 50 GTCTACGACGGTGACCTGCCGAGGGACATGTCCTGATCACCGCCGCCA 3550
 V Y D G D L P E G H V L I T A A H
 CCCCCGACGCCGAGGGACATACCCACCCGCCACACCCGCCACCC 3600
 P D D P E D I P T R A H T R A T
 GCGTCCTGACCGCCCTGCAACACCCACCTCACCAACCGACCAACCCCTC 3650
 55 R V L T A L Q H H L T T T D H T L
 ATCGTCCACACCAACCCACCGACCCCGCCGGCGCCACCGTCACTGGCCTCAC 3700
 I V H T T T D P A G A T V T G L T
 CGCGACCCGCCAGAACGAAACACCCACCGCATCCGCTCATCGAAACCG 3750
 R T A Q N E H P H R I R L I E T
 60 ACCACCCCCACACCCCCCTCCCCCTGGCCCAACTCGCCACCTCGACCAAC 3800

D H P H T P L P L A Q L A T L D H
 CCCCCACCTCCGGCCTCACCCACCCACACCCCTCCACCACCCCCCACCTCACCCC 3850
 P H L R L T H H T L H H P H L T P
 5 CTCACACACCACCCACCCACCCACCCACCCCTCAACCCCCAACACG 3900
 L H T T T P P T T T P L N P E H
 CCATCATCATCACCGGGCTCCGGCACCCCTCGCCGGCATCCTCGCCCGC 3950
 A I I I T G G S G T L A G F L A R
 CACCTGAACCACCCCCACACCTACCTCCTCCGCACCCCACCCCCCGA 4000
 H L N H P H T Y L L S R T P P P D
 10 CGCCACCCCCGGCACCCACCTCCCGTGCACGTCGGCACCCCCACCAAC 4050
 A T P G T H L P C D V G D P H Q
 TCGCCACCACCCCTCACCCACATCCCCAACCCCTCACCGCCATCTTCCAC 4100
 L A T T L T H I P Q P L T A I F H
 15 ACCGCCGCCACCCCTCGACGACGGCATTCCACGCCCTCACCCCCGACCG 4150
 T A A T L D D G I L H A L T P D R
 CCTCACCAACCGTCTCCACCCCCAAAGCCAACGCCGCCTGGCACCTGCACC 4200
 L T T V L H P K A N A A A W H L H
 ACCTCACCCAAAACCAACCCCTCACCCACTTCGTCTACTCCAGCGCC 4250
 20 H L T Q N Q P L T H F V L Y S S A
 GCCGCCGTCTCGGCAGCCCCGACAAGGAAACTACGCCGCCAACGC 4300
 A A V L G S P G Q G N Y A A A N A
 CTTCCCTCGACGCCCTGCCACCCACCGGCCACCCCTGGCCAACCCGCCA 4350
 F L D A L A T H R H T L G Q P A
 25 CCTCCATGCCCTGGGCATGTGGCACACCACAGCACCCCTCACCGGACAA 4400
 T S I A W G M W H T T S T L T G Q
 CTCGACGACGCCGACCCGGACCGCATCCGGCGGGTTCTCCCGAT 4450
 L D D A D R D R I R R G G F L P I
 CACGGACGACGAGGGCATGGGATGCAT
 30 T D D E G

The *AvrII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 13 (specific for methylmalonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

35 AGATCTGGCAGCTCGCCGAAGCGCTGCTGACGCTCGTCCGGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 GCCGCCGTGCTCGGCCACGTGGGTGGCGAGGACATCCCCGGCACGGCGGC 100
 A A V L G H V G G E D I P A T A A
 STCAAGGACCTCGGCATCGACTCGCTACCGCGGTCCAGCTGCACAG 150
 F K D L G I D S L T A V Q L R N
 CCCTCACCGAGGCACCGGTGTGCGGCTGAACGCCACGGCGGTCTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTCGCCGGGAAGCTCGCGACGAACGTGACCGG 250
 F P T F H V L A G K L G D E L T G
 40 45 CACCCGGCGCCCGCTCGTGCCTGGGACCGCGGCCACGGCGGTGCGCACG 300
 T R A P V V P R T A A T A G A H
 ACGAGCCGCTGGCAGTCGTGGGAATGGCCTGCCGCTGCCGGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 50 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCGCGCTGGGACGTCGACGCCATACGACC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCCGACGCCATCGCAAGACCTCGTCCGGCACGGTGGCTTCCTC 500
 55 P D P D A I G F T F V R H G G F L
 ACACGGCGCAGAGGCTTCGACGCCGCTTCGGCATCAGCCCGCGCA 550
 T G A T G F D A A F F G I S P R E
 GGCCTCGCGATGGACCCCGCAGCAGCGGGTGTCTGGAGACGTCGTGGG 600

A L A M D P Q Q R V L I E T S W
 AGGCCTTCGAAAGCCCGGCATCACCCGGACTCGACCCGGCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 AGCGCGCTGTTCTCGCGCTCTACGGTTACGGCACCCTCGGA 700
 5 T G V F T G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGACCGGCTCGCAGACCAAGTGTGCTCTCGGCC 750
 T D G F S A T G S Q T S V I S G
 GGCTGTGTTACTTCTACGGTCTGGAGGGCCGGCTACGGTACACG 800
 R L S Y F Y G L E G P A V T V D T
 10 CCGTGTGTTCTCGCTCGCTGGTGGCGCTGCACCAGGCCGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGCTGCCCTGGTCGGCGCTACGGTGATGGCGT 900
 S G E C S L A L V G G V T V M A
 CTCCCAGGCGGCTCGTGGAGTTCTCCGGCAGCGCGCCCTCGCGCCGGAC 950
 15 S P G G F V E F S R Q R G L A P D
 GGCGGGCGAAGGCCTCGCGCGGGTGCGGACGGCACGAGCTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 CGGTGCCGGTGTGCTGATCGTCAAGAGGGCTCTCCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 20 GTCACACCGTCTGGCGTCTCCGTGGTCAGCGCGCCCTCGCGCCGGAT 1100
 G H T V I A V V R G S A V N Q D G
 GCCTCCACGGGCTCTGGCGCCGAACGGGCCGTCGAGGAGCGGGTGAT 1150
 A S N G I S A P N G P S 2 E R V I
 CCGGCAGGCCCTGGCCAACGCCGGCTACCCCGGCCGACGTGGACGCCG 1200
 25 R Q A L A N A G L T P A D V D A
 TCGAGGCCACGGCACCGGCACCCAGGCTGGCGACCCCATCGAGGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 GCGGTACTGGCCACCTACGGACAGGAGCGCCACCCCCCTGCTCGTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 30 CTCGCTGAAGTCCAACATCGGCCACGCCAACGCCAGGCCGCTCCGGCGTC 1350
 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTGCAGGCCCTCCGGCACGGGGAGCTGCCGCCACG 1400
 G I I K M V Q A L R H G E L P P T
 35 CTGCACGCCACGAGCCGTCGCCGACGTGACTGGACGCCGGCGCCGT 1450
 L H A D E P S P H V D W T A G A V
 CGAACTGCTGACGTCGGCCGGCGTGGCCCGAGACCGACCGCTAGGC 1500
 E L L T S A R P W P E T D R P R
 GGGCGGGCGTGTGCTCTCGGAGTCAGCGCACCAACGCCACGTCATC 1550
 R A G V S S F G V S G T N A H V I
 40 CTGGAGAGCGCACCCCGCTCAGGCCGCGAGGAGGCCAGCCTGTTGA 1600
 L E S A P P A Q P A E E A Q P V E
 GACGCCCGTGGTGGCCCTCGGATGTGCTGCCGCTGGTATATCGGCCAAGA 1650
 T P V V A S D V L P L V I S A K
 CCCAGCCCGCCCTGACCGAACACGAAGACGGCTGCCGCCACCTGGCG 1700
 45 T Q P A L T E H E D R L R A Y L A
 GCGTCGCCGGGCGGGATACGGGCTGTGGCATCGACCGCTGGCGTGAC 1750
 A S P G A D I R A V A S T L A V T
 ACGGTGGTGTGCTGGAGCACGCCGTACTCTTGGAGATGACACCGTCA 1800
 R S V F E H R A V L L G D D T V
 50 CCGGCACCGCGGTGACCGACCCAGGATCGTGTGTTGCTTCCGGCGAG 1850
 T G T A V T D P R I V F V F P G Q
 GGGTGGCAGTGGCTGGGATGGCAGTGCAGTGCCTGGCGATCGTGGTGGT 1900
 G W Q W L G M G S A L R D S S V V
 GTTCGCCGAGCGGATGCCGAGTGTGCGGCCGTTGCCGAGTCGTGG 1950
 55 F A E R M A E C A A A L R E F V
 ACTGGGATCTGTTACGGTTCTGGATGATCCGGCGGTGGACCGGGTT 2000
 D W S L F T V L D D P A V V D R V
 GATGTGGTCCAGCCCGCTCTGGCGATGATGGTTCCCTGGCGCGGT 2050
 D V V Q P A S W A M M V S L A A V
 60 CTGGCAGGCCGCCGCTGCGGCCGGATGCCGATGGCCATTGCGAGG 2100

W Q A A G V R P D A V I G H S Q
 GTGAGATCGCCGCAGCTTGTGTGGCGGGTGCAGGTGCACTACGCCATGCC 2150
 G E I A A A C V A G A V S L R D A
 5 SCCCAGGATCGTACCTTCCGCAGCCAGGGATGCCCGGGCTGCCGG 2200
 A R I V T L R S Q A I A R G I A G
 CGGGGGCGCGATGCCATCCGTGCCCTGCCCGCGCAGGATGTGAGACTGG 2250
 R G A M A S V A L P A Q D V E L
 TCGACGGGGCCTGGATCGCCGCCACAACGGGCCGCCACCGTGTGATC 2300
 V D G A W I A A H N G P A S T V I
 10 GCGGGCACCCCGGAAGCGGTGACCATGTCCTCACCGCTCATGAGGCACA 2350
 A G T P E A V D H V L T A H E A Q
 AGGGGTGCGGGTGCGGGATCACCGTCGACTATGCCTCGCACACCCGC 2400
 G V R V R R I T V D Y A S H T P
 15 ACGTCGAGCTGATCCCGACGAACACTCGACATCACTAGCGACAGCAGC 2450
 H V E L I R D E L L D P T S D S S
 TCGCAGACCCCGCTCGTGCCTGGCTGACCGTGGACGGCACCTGGGT 2500
 S Q T P L V P W L S T V D G T W V
 CGACAGCCCGCTGGACGGGAGTACTGGTACCGGAACCTGCGTGAACCGG 2550
 D S P L D G E Y W Y R N L R E P
 20 TCGGTTTCCACCCCGCCGTAGCCAGTTGCAGGCCAGGGCGACACCGTG 2600
 V G F H P A V S Q L Q A Q G D T V
 TTCTCGAGGTGAGCCAGGCCAGGGTGTGTCAGGGATGGACGACGA 2650
 F V E V S A S P V L L Q A M D D D
 TGTCTCACGGTTCCCACGCTCGTCGTGACGACGGGACGCCACCGGA 2700
 25 V V T V A T L R R D D G D A T R
 TGCTCACCGCCCTGGCACAGGCCTATGTCACGGCTCACCGTCACTGG 2750
 M L T A L A Q A Y V H G V T V D W
 CCCGCCATCCTCGGACCCACCACAACCCGGTACTGGACCTCCGACCTA 2800
 P A I L G T T T R V L D L P T Y
 30 CGCCTTCCAACACCGGGTACTGGCTCGAGTCGGCACGCCCGGCCAT 2850
 A F Q H Q R Y W L E S A R P A A
 CCGACGCGGGCCACCCCGTGTGGCTCCGGTATGCCCTGCCGGTCG 2900
 S D A G H P V L G S G I A L A G S
 35 CCGGGCCGGTGTTCACGGGTTCCGTGCCGACCGGTGGGACCGCGCGT 2950
 P G R V F T G S V P T G A D R A V
 GTTCGTCGCCGAGCTGGCGCTGGCCGCCGGACGCGGTGACTGCCA 3000
 F V A E L A A A D A V D C A
 CGGTCGAGCGGCTCGACATCGCTCCGTGCCGGCCGGCATGGC 3050
 T V E R L D I A S V P G R P G H G
 40 CGGACGACCGTACAGACCTGGGTGACGAGCCGGACGACGGCGCG 3100
 R T T V Q T W V D E P A D D D G R R
 CCGGTTCACCGTGCACACCCGCACCGCGACGCCCGTGGACGCGTGCACG 3150
 R F T V H T R T G D A P W T L H
 CCGAGGGGGTGTGCCCTCCATGGCACGGCCCTGCCCGATGCCGCGAC 3200
 45 A E G V L R P H G T A L P D A A D
 GCCGAGTGGCCCCCACCGGGCGCGGTGCCCGGACGGCTGCCGGTGT 3250
 A E W P P P G A V P A D G L P G V
 GTGGCGCCGGGGGACCAAGGTCTTCGCCGAGGCCGAGGTGGACGGACCGG 3300
 W R R G D Q V F A E A E V D G P
 50 ACGGTTTCTGGTGCACCCCGACCTGCTCGACGCCGTCTCCGCCGGTC 3350
 D G F V V H P D L L D A V F S A V
 GGCACGGAAGCCGCCAGCCGGGATGGCGCGACCTGACGGTGCACGC 3400
 G D G S R Q P A G W R D L T V H A
 GTCGGACGCCACCGTACTGCCGCCCTGCCACCCGGCGACCGACGGAG 3450
 55 S D A T V L R A C L T R R T D G
 CCATGGGATTGCCGCCCTCGACGGCGCCGGCTGCCGGTACTCACCGCG 3500
 A M G F A A F D G A G L P V L T A
 GAGGGCGGTGACGCTGCCGGAGGTGGCGTACCGTCCGCCGAGGGAGTC 3550
 E A V T L R E V A S P S G S E E S
 60 CGACGCCCTGCAACCGTTGGAGTGGCTCGCGGTGCCGAGGCCGGTACG 3600

D G L H R L E W L A V A E A V Y
 ACGGTGACCTGCCCGAGGGACATGTCCTGATCACCGCCGCCACCCCGAC 3650
 D G D L P E G H V L I T A A H P D
 GACCCCGAGGGACATACCCACCCGCCACACCCGCCACCCCGTCT 3700
 5 C P E D I P T R A H T R A T R V L
 GACCGCCCTGCAACACCACCTCACCAACCACCGGACCAACCCCTCATCGTCC 3750
 T A L Q H H L T T D H T L I V
 ACACCAACCAACCGACCCCCCGGGCCACCGTCAACCGGCCTCACCCGCACC 3800
 10 H T T T D P A G A T V T G I T R T
 CCCCCAGAACGAACACCCCCACCGCATTCCGCTCATCGAAACCGGACCAACCC 3850
 A Q N E H P H R I R L I E T D H P
 CCACACCCCCCTCCCCCTGGCCCAACTCGCCACCCCTGACCACCCCCACC 3900
 H T P L P L A Q L A T L D H P H
 TCCGCCTCACCCACCAACACCCCTCACCAACCCCCACCTCACCCCCCTCCAC 3950
 15 L R L T H H T L H H P H L T P L H
 ACCACCACCCCCACCCACCCACCAACCCCCCTCAACCCCGAACACGCCATCAT 4000
 T T T P P T T T P L N P E H A I I
 CATCACCGGGCGCTCCGGCACCCCTCGCCGGCATCTCGCCGCCACCTGA 4050
 I T G G S G T L A G I L A R H L
 20 ACCACCCCCACACCTACCTCCTCTCCGCACCCACCCCCCGACGCCACC 4100
 N H P H T Y L L S R T P P P D A T
 CCCGGCACCCACCTCCCCCTGCGACGTCGGGACCCCCACCAACTGCCAC 4150
 P G T H L P C D V G D P H Q L A T
 CACCCCTCACCCACATCCCCAACCCCTCACCGCCATCTTCACACCGCCG 4200
 25 T L T H I P Q P L T A I F H T A
 CCACCCCTGACGACGGCATCCTCACGCCCTCACCCCCGACCGCCCTCAC 4250
 A T L D D G I L H A L T P D R L T
 ACCGTCCCTCCACCCCCAAAGCCAACGCCGCTGGCACCTGCACCACTCAC 4300
 T V L H P K A N A A A W H L H H L T
 30 CCAAAACCAACCCCTCACCCACTCGTCCTACTCCAGCGCCGCCGCG 4350
 Q N Q P L T H F V L Y S S A A A
 TCTCTGGCAGCCCCGGACAAGGAAACTACGCCGCCAACGCCCTCCTC 4400
 V L G S P G Q G N Y A A A A N A F L
 GACGCCCTGCCACCCACCGCCACACCCCTCGGCCAACCGCCACCTCCAT 4450
 35 D A L A T H R H T L G Q P A T S I
 CGCCTGGGGCATGTGGCACACCACAGCACCCCTCACCGGACAACCTGACG 4500
 A W G M W H T T S T L T G Q L D
 ACGCCGACCGGGACCCGATCCGCCGCCGGCTTCTCCGATCACGGAC 4550
 40 D A D R D R I R R G G F L P I T D
 GACGAGGGCATGGGATGCAT
 D E G

The *NheII-XhoI* restriction fragment that encodes module 8 of the FK-520 PKS with the endogenous AT domain replaced by the AT domain of module 12 (specific for malonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

AGATCTGGCAGCTGCCGAAGCGCTGCTGACGCTCGTCCGGAGAGGCC 50
 Q L A E A L L T L V R E S T
 50 G C C C C G T G C T C G G C C A C G T G G G T G G C G A G G A C A T C C C C G C G A C G G C G G C 100
 A A V L G H V G G E D I P A T A A
 G T T C A A G G A C C T C G G C A T C G A C T C G C T C A C C G C G G T C A G C T G C G C A A C G 150
 F K D L G I D S L T A V Q L R N
 C C C T C A C C G A G G G C A C C G G T G T G C G G C T G A A C G C C A C G G C G G T C T T C G A C 200
 A L T E A T S V R L N A T A V F D
 T T C C C G A C C C C G C A C G T G C T C G C C G G G A A G C T C G G C G A C G A A C T G A C C G G 250
 F P T P H V L A G K L G D E L T G
 C A C C C G C G C C C C G T G C T G C C C G G A C C G C G G C C A C G G C C G G T G C G C A C G 300

T R A F V V P R T A A T A G A H
 ACGAGCCGCTGGCGATCGTGGGAATGGCTGCGGCTGCCGGCTGGCGGGTC 350
 D E P L A I V G M A C R L P G G V
 CGGTCAACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 5 A S P E E L W H L V A S G T D A I
 CACGGAGTTCCCGACGGACCGCGGCTGGGACGTCGACCGCAGTCGAC 450
 T E F P T D R G W D V D A I Y D
 CGGACCCGACGCGATCGGCAAGACCTCGTCCGGCACGGTGGCTTCCTC 500
 F D P D A I G K T F V R H G G F L
 10 ACCGGCGCAGAGCTTCGACGGCGCTTCTCGGCATCAGCCCGCGA 550
 T G A T G F D A A F F G I S P R E
 GCCCCTCGCGATGGACCCCGCAGCAGCGGGTGCCTGGAGACGTCGTGGG 600
 A L A M D P Q Q R V L L E T S W
 AGGCGTTCGAAAGCGCCGGCATCACCCCGACTCGACCCCGGGCAGCGAC 650
 15 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCCTCTCCTACGGTTACGGCACCGGTGCGGA 700
 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGCAGCCGCTCGCAGACCAGTGTGCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 20 GGCTGTCGTAATTCTACGGTCTGGAGGGTCCGGCGGTACGGTCGACACG 800
 R L S Y F Y G L E G P A V T V D T
 CGGTGTTCTCGTCGCTGGTGGCGCTGCACCAGGCCGGCAGTCGCTGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGTCGCTCGCCCTGGTCGGCGGTACGGTGATGGCGT 900
 25 S G E C S L A L V G V T V M A
 CTCCCAGGCGCTTCTGGAGTTCTCCGGCAGCGCGCCCTCGCCGGAC 950
 S P G G F V E F S R Q R G L A P D
 GCCCGGGCGAAGGCCTTCGGCGGGTGCGGACGGCACGAGCTCGCCGA 1000
 30 G R A K A F G A G A D G T S F A E
 GGCTGCCGGTGTGCTGATCGTCGAGAGGCTCTCCGACGCCAACGCAACG 1050
 G A G V L I V E R L S D A E R N
 GTACACACCGTCTGGCGGTCTCCGTGGTTCGGCGGTCAACCAGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 35 GCCTCCAACGGCTGTCGGCGCCAACGGCGCTCGCAGGAGCGGGTGA 1150
 A S N G L S A P N G P S Q E R V I
 CCGGCAGGCCCTGGCAAACGCCGGCTACCCCGGGACGTGGACGCCG 1200
 R Q A L A N A G L T P A D V D A
 TCGAGGCCAACGGCACCGGACCCAGGCTGGCGACCCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 40 GCGGTACTGCCAACCTACGGACAGGAGCGCGCCACCCCCCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 CTCGCTGAAGTCAAACATCGGCCACGCCAACGGCGCTCCGGCGTCCGGCG 1350
 S L K S N I G H A Q A A S G V A
 GCATCAAGATGGTGCAGGCCCTCCGGCACGGGGAGCTGCCGCCACG 1400
 45 S I I K M V Q A L R H G E L P P T
 CTGCACGCCACGCCGTCGCCACGTGACTGGACGGCCGGCGCCGT 1450
 I H A D E P S P H V D W T A G A V
 CGACTGCTGACGTCGCCGGCGTGGCCGAGACCGACCGGCCACGGC 1500
 E L L T S A R P W P E T D R P R
 50 CTGCCGCCGTCTCTCGTGGGTGAGCGGCACCAACGCCACGTCACTC 1550
 R A A V S S F G V S G T N A H V I
 CTGSAGGCCGGACCGTAACGGAGACGCCCGGGCATGCCCTCGGTGA 1600
 L E A G P V T E T P A A S P S G D
 CCTTCCCTGCTGGTGTGGCACGCTCACCGGAAGCGCTCGACGAGCAGA 1650
 55 L P L L V S A R S P E A L D E Q
 TCCGCCGACTGCCCTACCTGGACACCAACCCCGGACGTCGACCGGGTG 1700
 T R R L R A Y L D T T P D V D R V
 CCTGCTGGCACAGACGCTGGCCGGCGACACACTTCCCCACCCCGCCGT 1750
 A V A I Q T L A R R T H F A H R A V
 60 GCTGCTGGTGAACCGTCATCACCAACCCCCCGCGGACCGGGCCGACG 1800

L L G D T V I T T P P A D R P D
 AACTCGTCTCGTCTACTCCGGCCAGGGCACCCAGCATCCCGCATGGGC 1850
 E L V F V Y S G Q G T Q H P A M G
 GAGCAGCTAGCGCCCGCGTTCCCCGTCTCGCGCCTGCATCAGCAGGT 1900
 5 E Q L A A A F P V F A R I H Q Q V
 GTGGGACCTGCTCGATGTGCCCGATCTGGAGGTGAAACGAGACCGGTTACG 1950
 W D L I D V P D L E V N E T G Y
 CCCAGCCGGCCCTGTTCGCAATGCAGGTGGCTCTGTTGGGCTGCTGGAA 2000
 10 A Q P A I F A M Q V A L F G I L E
 TCGTGGGCTGACGACCCGACGCCGTATCGGGCCATTGGTGGGTGAGCT 2050
 S W G V R P D A V I G H S V S E L
 TGCGGCTCGTATGTGTCGGGGTGTGGTCGTTGGAGGATGCCTGCACTT 2100
 A A A Y V S G V W S L E D A C T
 TGGTGTGGCGCGGGCTCGTCTGATGCAGGCTCTGCCCGCGGTGGGTG 2150
 15 L V S A R A R L M Q A L P A S G V
 ATGGTCGCTGTCCGGTCTCGGAGGATGAGGCCCGGGCGTGTGGGTGA 2200
 M V A V P V S E D E A R A V I G E
 GGGTGTGGAGATCGCCCGGGTCAACGGCCCGTGTGGTTCTCTCCG 2250
 G V E I A A V N G P S S V V L S
 20 GTGATGAGGCCGCCGTGCTGCAGGCCGCCAGGGCTGGGAAGTGGACG 2300
 G D E A A V L Q A A E G I G K W T
 CGGCTGGCGACCAGCCACGCCGTCCATTCCGCCCTATGGAACCCATGCT 2350
 R L A T S H A F H S A R M E P M L
 GGAGGAGTTCCGGCGGTGCCGAAGGCCGTGACCTACCGGACGCCCGAGG 2400
 25 E E F R A V A E G L T Y R T P Q
 TCTCCATGGCCCTGGTGATCAGGTGACCAACCGCTGAGTACTGGGTGCGG 2450
 V S M A V G D Q V T T A E Y W V R
 CAGGTCCGGGACACGGTCCGGTCCGGCAGCAGGTGGCCTCGTACGAGGA 2500
 Q V R D T V R F G E Q V A S Y E D
 30 CGCCGTGTTCGTCGAGCTGGTGCCGACCGGTCACTGCCCGCTGGTCG 2550
 A V F V E L G A D R S I A R L V
 ACGGTGTGGCGATGCTGCACGGCGACCACGAAATCCAGGCCGATCGGC 2600
 D G V A M L H G D H E I Q A A I G
 GCCCTGGCCCACCTGTATGTCACCGCGTCACGGTCCACTGGCCCGCGCT 2650
 35 A L A H L Y V N G V T Y D W P A L
 CCTGGGCGATGCTCCGGAACACGGGTGCTGGACCTCCGACATACGCGCT 2700
 L G D A P A T R V L D I P T Y A
 TCCAGCACCGCGTACTGGCTCGAGTCGGCACGCCCGCCGATCCGAC 2750
 F Q H Q R Y W L E S A R P A A S D
 40 GCAGGGCCACCCCGTGCTGGGCTCCGGTATGCCCTCGCCGGTCCCGGG 2800
 A G H P V L G S G I A L A G S P G
 CGGGTGTTCACGGGTTCCCGTCCGGACCGGTGGCGACCGCGCGGTGTTCG 2850
 R V F T G S V P T G A D R A V F
 TCGCCGAGCTGGCGCTGGCCGCCGGACGCCGTGACTGCGCCACGGTC 2900
 45 V A E L A L A A A D A V D C A T V
 GAGCGGCTGACATCGCCTCCGTGCCCGGCCGGCCGACATGGCCGGAC 2950
 E R L D I A S V P G R P G H G R T
 GACCGTACAGACCTGGGTGACGAGGCCGGCGACGACGGCCGGCGCGT 3000
 T V Q T W V D E P A D C G R P R
 50 TCACCGTGCACACCCGACCGGCACGCCCGTGGACCGCTGCAACCCGAG 3050
 F T V H T R T G D A P W I L H A E
 GGGGTGCTGGCCCCCATGGCACGGCCCTGCCCGATGCCGGACGCCGA 3100
 G V L R P H G T A L P D A A D A E
 GTGGCCCGACCCGGCGCGGTGCCCGCGACGGGTGCGGGTGTGGC 3150
 55 W P P P G A V P A D G I P G V W
 GCCGGGGGGACCAAGGTCTTCGCCGAGGCCGGAGGTGGACGGACCGGACGGT 3200
 R R G D Q V F A E A E V D G P D G
 TTGCTGGTGCACCCCGACCTGCTCGACGCCGGTCTTGTGGCGATGGCGA 3250
 F V V H P D L L D A V F S A V G D
 CGGAAGCCGCCAGCCGGGATGGCGCACCTGACGGTGCACCGTCGG 3300

S S R Q P A G W R D L T V H A S
 ACGCCACCGTACTGCGCGCTGCCTCACCCGGCGCACCGACGGAGGCCATG 3350
 D A T V L R A C L T R R T D G A M
 5 GATTCCCGCGCTTCGACGGCGCCGGCTGCCGGTACTCACCCGGAGGC 3400
 G F A A F D G A G L P V L T A E A
 GGTGACGCTGCGGGAGGTGGCGTACCGTCCGGCTCCGAGGAGTCGGACG 3450
 V T L R E V A S P S G S E E S D
 GCGCTGCACCGGTTGGAGTGGCTCGCGTCCGGCTACGACGGT 3500
 10 G L H R L E W L A V A E A V Y D G
 GACCTGCCCGAGGGACATGTCCGTGATCACCGCGCCACCCGACGACCC 3550
 D L P E G H V L I T A A H P D D P
 CGAGGACATACCCACCCGCGCCACACCCGCGCCACCCGCGTCTGACCG 3600
 E D I P T R A H T R A T R V L T
 15 CCCTGCAACACCCACCTCACCAACCACCGGACACACCCCTCATCGTCCACACC 3650
 A L Q H H L T T T D H T L I V H T
 ACCACCGACCCCGCCGGCGCCACCGTACCGGGCTCACCGCACCGCC 3700
 T T D P A G A T V T G L T R T A Q
 GAACGAAACACCCACCCGACATCCGCTCATCGAAACCGGACCAACCC 3750
 20 N E H P H R I R L I E T D H P H
 CCCCCCTCCCCCTGGCCCAACTCGCCACCCCTCGACCACCCCCCACCTCCGC 3800
 T P L P L A Q L A T L D H P H L R
 CTCACCCACACACCCCTCACCAACCCCCCACCTCACCCCCCTCGACACAC 3850
 L T H H T L H H P H L T P L H T T
 CACCCCCACCCACCAACCCCCCTCAACCCCGAACACGCCATCATCA 3900
 25 T P P T T T P L N P E H A I I I
 CCGGCGGCTCCGGCACCCCTCGCCGGCATCTCGCCGCCACCTGAACAC 3950
 T G G S G T L A G I L A R H L N H
 CCCCCACACCTACCTCCCTCCGGCACCCCCACCCCCCGACGCCACCCCCGG 4000
 P H T Y L L S R T P P P D A T P G
 30 CACCCACCTCCCCCTCGACGTCGGCGACCCCCACCAACTCGCCACCAAC 4050
 T H L P C D V G D P H Q L A T T
 TCACCCACATCCCCAACCCCTCACGCCATCTTCCACACCGCCGCCACC 4100
 L T H I P Q P L T A I F H T A A T
 CTCGACGACGGCATCCTCCACGCCCTCACCCCCGACGCCACCCCGT 4150
 35 L D D G I L H A L T P D R L T T V
 CCTCCACCCCCAAAGCCAACGCCGCGCTGGCACCTGCACCACTCACCC 4200
 L H P K A N A A W H L H H L T Q
 ACCAACCCCTCACCCACTTCGTCCCTACTCCAGGCCGCCGCGCTC 4250
 N Q P L T H F V L Y S S A A A V L
 40 GGCAGCCCCGGACAAGGAAACTACGCCGCCAACGCCCTCGACGC 4300
 G S P G Q G N Y A A A N A F L D A
 CCTCGCCACCCACCGCCACACCCCTCGGCCAACCGCCACCTCCATCGCCT 4350
 L A T H R H T L G Q P A T S I A
 GGGGCATGTGGCACACCACAGCACCCCTCACCGGACAACCTCGACGCC 4400
 45 W G M W H T T S T L T G Q L D D A
 GACCGGGACCGCATCCGCCGCCGGTTCTCCCGATCACGGACGACGA 4450
 D R D R I R R G G F L P I T D D E
 GGGCATGGGGATGCAT
 G

50 The *NheII-XbaI* restriction fragment that encodes module 8 of the FK-520 PKS with

the endogenous AT domain replaced by the AT domain of module 13 (specific for methylmalonyl CoA) of the rapamycin PKS has the DNA sequence and encodes the amino acid sequence shown below.

55 AGATCTGGCAGCTGCCAAGCGCTGCTGACGCTCGTCCGGAGAGCACC 50
 Q L A E A L L T L V R E S T
 GCGCCCGTGCCTGGCACGTGGTGGCGAGGACATCCCCGCGACGGCGC 100

A A V L G H V G G E D I P A T A A
 GTTCAAGGACCTCGGCATCGACTCGCTCACCGCGGTCAGCTGCCAACG 100
 F K D L G I D S L T A V Q L R N
 5 CCGTCAACGGACGGCGACCGGTGTGGCTGAAACGCGACCGCGGCTTCGAC 200
 A L T E A T G V R L N A T A V F D
 TTCCCGACCCCGCACGTGCTCGCCGGAAAGCTCGGCACGAACCGGG 250
 F P T P H V L A G K L G D E L T G
 CACCCCGCGCGCCCGTGTGCCCCGGACCGCGGCCACGGCCGTGCGCACG 300
 T R A P V V P R T A A T A G A H
 10 ACGAGCCGCTCGCGATCGTGGAAATGGCCTGCCGGCTGCCGGGGTC 350
 D E P L A I V G M A C R L P G G V
 GCGTCACCCGAGGAGCTGTGGCACCTCGTGGCATCCGGCACCGACGCCAT 400
 A S P E E L W H I V A S G T D A I
 CACGGAGTTCCCACGGACCGCGCTGGGACGTCGACCGCGATCTACGACC 450
 15 T E F P T D R G W D V D A I Y D
 CGGACCCGACCGCGATCGGAAAGACCTTCGTCCGGCACGGTGGCTTCCTC 500
 P D P D A I G K T F V R H G G F L
 ACCGGCGCGACAGGCTTCGACCGGGCTTCGGCATCAGCCCGCGCA 550
 T G A T G F D A A F F G I S P R E
 20 GGCCTCGCGATGGACCCGACAGCAGCGGGTGCTCTGGAGACGTCGTGGG 600
 A L A M D P O Q R V L L E T S W
 AGGCCTTCGAAAGCCCCGGCATCACCCCGACTCGACCCCGCGCAGCGAC 650
 E A F E S A G I T P D S T R G S D
 ACCGGCGTGTTCGTCGGCGCCTCTCTACGGTTACGGCACCGGTGCGGA 700
 25 T G V F V G A F S Y G Y G T G A D
 CACCGACGGCTTCGGCGCGACCGGCTCGCAGACAGCAGTGTGCTCTCCGGCC 750
 T D G F G A T G S Q T S V L S G
 GGCTGTCGTACTTCTACGGCTCTGGAGGGTCCGGCGGTACGGTCGACACG 800
 R L S Y F Y G L E G P A V T V D T
 30 GCGTGTTCGTCGTGCTGGCTGCACCAGGCCGGCAGTCGCTCGCG 850
 A C S S S L V A L H Q A G Q S L R
 CTCCGGCGAATGCTCGTCGCCCTGGTCGGCGCGTCACGGTATGGCGT 900
 S G E C S L A L V G G V T V M A
 CTCCGGCGGCTTCGAGGTTCTCCGGCAGCGCGGCCCTCGCGCCGGAC 950
 35 S P G G F V E F S R Q R G L A P D
 GGCCTGGCGAAGGCCTCGGCGGGTGCAGCGCACGGCACGAGCTCGCCGA 1000
 G R A K A F G A G A D G T S F A E
 GGGTCCGGTGTGCTGATCGTCAGAGGGCTCTCGACGCCAACGCAACG 1050
 S A G V L I V E R L S D A E R N
 40 GTCATCACCGCTCTGGCGGTCTGGTTCTGGCTGCAACCAGGATGGT 1100
 G H T V L A V V R G S A V N Q D G
 GCTCCAACGGGCTGTCGGCGCCAACGGGGCGTCGAGGAGCGGGTGT 1150
 A S N G L S A P N G P S Q E R V I
 CCGGCAGGCCCTGGCCAACGCCGGCTCACCCCGCGACGTGGACGCC 1200
 45 R Q A L A N A G L T P A D V D A
 TCGAGGCCACGGCACCCGGCACAGGCTGGCGACCCATCGAGGCACAG 1250
 V E A H G T G T R L G D P I E A Q
 CCGGTACTGGCCACCTACGGACAGGAGCGGCCACCCCCCTGCTGGG 1300
 A V L A T Y G Q E R A T P L L L G
 50 CTCGCTGAAGTCCAACATCGGCCACGCCAACGGCGCGTCCGGCGTCCG 1350
 S L K S N I G H A Q A A S G V A
 GCATCATCAAGATGGTCAGGCCCTCGGCACGGGAGCTGCCGCCGACG 1400
 S I I K M V Q A L R H G E L P P T
 CTGCACGCCGACGAGCCCTCGCCCGACGTGCACTGCACGCCGGCGCCGT 1450
 55 L H A D E P S P H V D W T A G A V
 CGAAGTGTGACGTCGGCCGGCGTGGCCCGAGACCGACCGGCCACGGC 1500
 F L L T S A P P W P E T D P P R
 GTCGGCGCTGTCGGGGCTGAGCGGCCACCGCCGACGTGATC 1550
 R A A V S S F G V S G T N A H V I
 GTCGAGGCCGACCGGTAACGGAGACGCCGGCATCGCCTCCGGTGA 1600

L E A G P V T E T P A A S P S G D
 CCTTCCCCCTGCTGGTGTGGCACGCTCACCGGAAGCGCTCGACCGAGCAGA 1650
 L P L L V S A R S P E A L D E Q
 TCCGCCGACTGCGCGCTACCTGGACACCACCCGGACGTCGACCGGGTG 1700
 5 I R R L R A Y L D T T P D V D R V
 GCCGTGGCACAGACGCTGGCCGGCGCACACACTTCGCCAACCGCGCCGT 1750
 A V A Q T L A R R T H F A H R A V
 GCTGCTCGGTGACACCGTCATCACCAACACCCCCCGCGGACCGGGCCGACG 1800
 L L G D T V I T T P P A D R P D
 10 AACCTCGTCTTCGTCTACTCCGGCCAGGGCACCCAGCATCCCGCATGGGC 1850
 E L V F V Y S G Q G T Q H P A M G
 GAGCAGCTAGCCGATTGTCGGTGGTGGTCGCCGAGCGGATGGCCGAGTG 1900
 E Q L A D S S V V F A E R M A E C
 TCGGGCGGCGTTGCGCGAGTTCGTGGACTGGGATCTGTTCACGGTTCTGG 1950
 15 A A A L R E F V D W D L F T V L
 ATGATCCGGCGGTGGTGGACCGGGTTGATGTGGTCCAGCCGCTTCCTGG 2000
 D D P A V V D R V D V V Q P A S W
 GCGATGATGGTTCCCTGGCCGCGGTGTGGCAGGCGGCCGGTGTGCGGCC 2050
 A M M V S L A A V W Q A A G V R P
 20 GGATGCGGTGATCGGCCATTGCAAGGGTGAGATGCCGCAGCTTGTGTGG 2100
 D A V I G H S Q G E I A A A C V
 CGGGTGCGGTGTCACTACCGGATGCCGCCGGATCGTGCACCTGCGCAGC 2150
 A G A V S L R D A A R I V T L R S
 CAGGCGATGCCGCCGGGCTGGCGGGCGCGATGGCATCCGTCGC 2200
 25 Q A I A R G L A G R G A M A S V A
 CCTGCCCCGCAGGATGTCGAGCTGGTCGACGGGGCTGGATGCCGCC 2250
 L P A Q D V E L V D G A W I A A
 ACAACGGGCCCGCCTCACCGTGTGATCGCGGGCACCCCGGAAGCGGTGAC 2300
 H N G P A S T V I A G T P E A V D
 30 CATGTCCTCACCGCTCATGAGGCACAAGGGTGCGGGTGCAGGATCAC 2350
 H V L T A H E A Q G V R V R R I T
 CGTCGACTATGCCTCGCACACCCCGCACGTCGAGCTGATCCGCGACGAAC 2400
 V D Y A S H T P H V E L I R D E
 TACTCGACATCACTAGCGACAGCAGCTCGCAGACCCCGCTCGTGCCTGG 2450
 35 L L D I T S D S S S Q T P L V P W
 CTGTCGACCGTGGACGGCACCTGGTGTGACAGCCGCTGGACGGGAGTA 2500
 L S T V D G T W V D S P L D G E Y
 CTGGTACCGGAACCTGCGTGAACCGTGGTTCCACCCGCCGTCAGCC 2550
 W Y R N L R E P V G F H P A V S
 40 ACTTGCAAGGCCAGGGGACACCGTGGTCGAGGTGACGGCTGGTGC 2600
 Q I Q A Q G D T V F, V E V S A S P
 GTGTTGTTGCAAGCGATGGACGACGATGTCGTACGGTTGCCACGCTGCG 2650
 V L L Q A M D D D V V T V A T L R
 TCGTGACGACGGCGACGCCACCCGGATGTCACCGCCCTGGCACAGGCCT 2700
 45 R D D G D A T R M L T A L A Q A
 ATGTCACGGCGTCACCGTCGACTGGCCGCCATCTCGGCACCACACA 2750
 Y V H G V T V D W P A I L G T T T
 ACCCGGGTACTGGACCTTCGACCTACGCCCTCCAACACCAGCGGTACTG 2800
 T R V L D L P T Y A F Q H Q R Y W
 50 GCTCGAGTCGGCACGCCGGCGATCCGACGCCGGCACCCGTGCTGG 2850
 L E S A R P A A S D A G H P V L
 CCTCCCGTATCGCCCTGCCGGTGGCCGGCGGTGTTCACGGGTCC 2900
 G S G I A L A G S P G R V F T G S
 GTGCCGACCGCGTGGACCGCGCGGGTGTTCGTCGCCGAGCTGGCGCTGGC 2950
 55 Y P T G A D R A V F V A E L A L A
 CGCCGCCGACGCCGGTCCACTGCCACGGTCCAGCGGCTCGACATGCCCT 3000
 A A D A V D C A T V E F L E I A
 CGTCGCCCCCGGCCGCGGGCATGCCCGAACGACGCTACAGACCTGGGTG 3050
 S V P G R P G H G R T T V Q T W V
 GACGAGGCCGGACGACGGCGGCCGGTCAACCGTGCACACCCGCAC 3100

D E F A D D G R R R F T V H T R T
 CCGCGACCCCGCTGGACCGCTCACGCCGAGGGGTCTGGCCCGATG 3150
 G D A P W T L H A E G Y L P P H
 CGACGGCGCTCGCGATGGCGACGCCGAGTGGCGCCGGGGACCAAGGTCTT 3200
 5 G T A L P D A A D A E W P P P G A
 GTGCCCGCGACCGCTGCCGGTGTGTGGCGCCGGGGACCAAGGTCTT 3250
 V P A E G L P G V W R R G D Q V F
 CGCCGAGGGCGAGGTGGACGGACCGGACGGTTCTGTGGTGCACCCGACC 3300
 A E A E V D G P D G F V V H P D
 10 TGCTCGACCGCGCTTCTCCGGCTGGCGACGGAAGCCGCCAGCCGGCC 3350
 L L D A V F S A V G D G S R Q P A
 GGATGGCGCGACCTGACGGTGACCGCTGGACGCCACCGTACTGCGCGC 3400
 G W R D L T V H A S D A T V L R A
 CTGCCCTCACCGCGACCGAGGCCATGGGATTGCCGCCCTCGACG 3450
 15 C L T R R T D G A M G F A A F D
 CGCCCGGGCTGCGCGTACTCACCGCGAGGCCTGACGCTGCCGGAGGTG 3500
 G A G L P V L T A E A V T L R E V
 GCGTCACCGCTCCCGCTCCGAGGAGTCGGACGCCCTGACCCGTTGGAGTG 3550
 A S P S G S E E S D G L H R L E W
 20 GCTCGCGGTGCGCGAGGCCTGACGGTGACCTGCCCGAGGGACATG 3600
 L A V A E A V Y D G D E P E G H
 TCCTGATCACCGCGCCACCCCGACGACCCCGAGGACATACCGACCG 3650
 V L I T A A H P D D P E D I P T R
 GCCCACACCCCGCGCCACCCCGCGTCTGACCGCCCTGACACCCACCTCAC 3700
 25 A H T R A T R V L T A L Q H H L T
 CACCAACGACCAACACCCCTCATCGTCCACACCACCGACCCCGCCGGCG 3750
 T T D H T L I V H T T D P A G
 CCACCGTCACCGCGCTCACCGCACCGCCAGAACGAAACACCCCGCC 3800
 A T V T G L T R T A Q N E H P H R
 30 ATCCGCCTCATCGAAACCGACCAACCCCCCACACCCCCCTCCCCCTGGCCCA 3850
 I R L I E T D H P H T P L P L A Q
 ACTCGCCACCCCTCGACCAACCCCCCACCTCCGCCCTACCCACACCCCTCC 3900
 L A T L D H P H L R L T H H T L
 ACCACCCCCCACCTCACCCCCCTCACACCAACCCCCACCCACCCACC 3950
 35 H H P H L T P L H T T T P P T T T
 CCCCTCAACCCCGAACACGCCATCATCACCGGGCGCTCCGGCACCCCT 4000
 P L N P E H A I I I T G G S G T L
 CGCCGGCATCCTCGCCCGCACCTGAACCAACCCCCCACACCTACCTCCCT 4050
 A G I L A R H L N H P H T Y L L
 40 CGCGCACCCCCACCCCCCGACGCCACCCCCCGCACCCACCTCCCTGCGAC 4100
 S R T P F P D A T P G T H L P C D
 GTGGGGCACCCCCCACCAACTCGCCACCCACCTCACCCACATCCCCAAC 4150
 V G D P H Q L A T T L T H I P Q P
 CCTCACCGCCATCTCCACACCGCCGCCACCCCTCGACGACGGCATCCCTCC 4200
 45 L T A I F H T A A T L D D G I L
 ACGCCCTCACCCCCGACCGCCCTCACCAACCGCTCCACCCAAAGCCAAC 4250
 H A L T P D R L T T V L H P K A N
 GCCGCCTGGCACCTGCACCCACCTCACCCAAAACCAACCCCTCACCCACTT 4300
 A A W H L H H L T Q N Q P L T H F
 50 CGTCCCTACTCCAGCGCCGCCGTCTGGCAGCCCCGGACAGGAA 4350
 V L Y S S A A A V L G S P G Q G
 ACTACGCCGCCAACGCCCTCTCGACGCCCTGCCACCCACCGCCAC 4400
 N Y A A A N A F L D A L A T H R H
 ACCCTCGGCCAACCCGCCACCTCCATCGCCTGGGGCATGTGGCACACCAC 4450
 55 T L G Q P A T S I A W G M W H T T
 CAGCACCCCTCACCGGACAACCTCGACGACGCCGACCGGACCCGATCCGCC 4500
 S T L T G Q L D D A D R D P I R
 CGGGGGGTTTCTCCCGATCACCGACGACGCCGAGGGGATGGGGATCGAT
 R G G F L P I T D D E G

Phage KC515 DNA was prepared using the procedure described in *Genetic Manipulation of Streptomyces*, A Laboratory Manual, edited by D. Hopwood *et al.* A phage suspension prepared from 10 plates (100 mm) of confluent plaques of KC515 on *S. lividans* TK24 generally gave about 3 µg of phage DNA. The DNA was ligated to circularize at the 5 cos site, subsequently digested with restriction enzymes *Bam*HI and *Pst*I, and dephosphorylated with SAP.

Each module 8 cassette described above was excised with restriction enzymes *Bgl*II and *Nsi*I and ligated into the compatible *Bam*HI and *Pst*I sites of KC515 phage DNA prepared as described above. The ligation mixture containing KC515 and various cassettes 10 was transfected into protoplasts of *Streptomyces lividans* TK24 using the procedure described in *Genetic Manipulation of Streptomyces*, A Laboratory Manual edited by D. Hopwood *et al.* and overlaid with TK24 spores. After 16-24 hr. the plaques were restreaked on plates overlaid with TK24 spores. Single plaques were picked and resuspended in 200 µL of nutrient broth. Phage DNA was prepared by the boiling method (Hopwood *et al.*, 15 *supra*). The PCR with primers spanning the left and right boundaries of the recombinant phage was used to verify the correct phage had been isolated. In most cases, at least 80% of the plaques contained the expected insert. To confirm the presence of the resistance marker (thiostrepton), a spot test is used, as described in Lomovskaya *et al.* (1997), in which a plate with spots of phage is overlaid with mixture of spores of TK24 and phiC31 TK24 lysogen. 20 After overnight incubation, the plate is overlaid with antibiotic in soft agar. A working stock is made of all phage containing desired constructs.

Streptomyces hygroscopicus ATCC 14891 (see US Patent No. 3,244,592, issued 5 Apr 1966, incorporated herein by reference) mycelia were infected with the recombinant phage by mixing the spores and phage (1 x 10⁸ of each), and incubating on R2YE agar 25 (*Genetic Manipulation of Streptomyces*, A Laboratory Manual, edited by D. Hopwood *et al.*) at 30°C for 10 days. Recombinant clones were selected and plated on minimal medium containing thiostrepton (50 µg/ml) to select for the thiostrepton resistance-conferring gene. Primary thiostrepton resistant clones were isolated and purified through a second round of single colony isolation, as necessary. To obtain thiostrepton-sensitive revertants that 30 underwent a second recombination event to evict the phage genome, primary recombinants were propagated in liquid media for two to three days in the absence of thiostrepton and then spread on agar medium without thiostrepton to obtain spores. Spores were plated to obtain about 50 colonies per plate, and thiostrepton sensitive colonies were identified by

replica plating onto thiostrepton containing agar medium. The PCR was used to determine which of the thiostrepton sensitive colonies reverted to the wild type (reversal of the initial integration event), and which contain the desired AT swap at module 8 in the ATCC 14891-derived cells. The PCR primers used amplified either the KS/AT junction or the AT/DH junction of the wild-type and the desired recombinant strains. Fermentation of the recombinant strains, followed by isolation of the metabolites and analysis by LCMS, and NMR is used to characterize the novel polyketide compounds.

Example 2

10 Replacement of Methoxyl with Hydrogen or Methyl at C-13 of FK-506

The present invention also provides the 13-desmethoxy derivatives of FK-506 and the novel PKS enzymes that produce them. A variety of *Streptomyces* strains that produce FK-506 are known in the art, including *S. tsukubaensis* No. 9993 (FERM BP-927), described in U.S. Patent No. 5,624,852, incorporated herein by reference; *S. hygroscopicus* subsp. *yakushimaensis* No. 7238, described in U.S. patent No. 4,894,366, incorporated herein by reference; *S. sp.* MA6858 (ATCC 55098), described in U.S. Patent Nos. 5,116,756, incorporated herein by reference; and *S. sp.* MA 6548, described in Motamedi *et al.*, 1998, "The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK-506," *Eur. J. Biochem.* 256: 528-534, and Motamedi *et al.*, 1997, "Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK-506," *Eur. J. Biochem.* 244: 74-80, each of which is incorporated herein by reference.

The complete sequence of the FK-506 gene cluster from *Streptomyces* sp. MA6548 is known, and the sequences of the corresponding gene clusters from other FK-506-producing organisms is highly homologous thereto. The novel FK-506 recombinant gene clusters of the present invention differ from the naturally occurring gene clusters in that the AT domain of module 8 of the naturally occurring PKSs is replaced by an AT domain specific for malonyl CoA or methylmalonyl CoA. These AT domain replacements are made at the DNA level, following the methodology described in Example 1.

30 The naturally occurring module 8 sequence for the MA6548 strain is shown below, followed by the illustrative hybrid module 8 sequences for the MA6548 strains.

35 5' G C A T G C G G C T G T A C G G A G G C G G C A C G G C G C A C C G G A A G T C C C G T G C T G G T C 50
 M R L Y E A A R R T G S P V V V
 G C G G C C G C G C T C G A C C G A C G C G C C G G A C G T G C C G C T G C T G C G C G G G C T G C G 100
 A A A L D D A P D V P L L R G L R

5' 3' CGTACGACCGTCCGGCGTCCGCCGCTCCGGGAAAGCTCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCGCCGACGCCCTCCCTCGCGTTCG 200
 R S P W C P T T S A P T P R S R S
 TCCGGAAACAGCACCCGCCACCGTGTCCGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CGGGGACGACGACGTTCAAGGAACCTCGGCATCGACTCGCTCACCGCGG 300
 P A T T T F K E L G I D S L T A
 TCCAGCTGCACACGCCGCTGACCACGGCGACCGGGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTTCCGACGCCGCCGCGCTCGCCGCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCGCCGCGCCGTCGCGGCCCGGACCGCGGCCA 450
 D E L A G T R A P V A A R T A A
 CGCGGCCGCGCACGACGAAACCGCTGGCGATCGTGGCATGGCCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGCGGGGTCGCGTCGCCACAGGAGCTGTGGCGTCTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 CGGCACCCGACGCCATCACGGAGTTCCCGCGGACCGCGGCTGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCCGACGCGATCGGAAGACCTTCGTCGG 650
 S A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCTCGACGGTGCACCGGCTTCGACGCCGCTTCGG 700
 H G G - F L D G A T G F D A A F F G
 25' 5' 25 5' GATCAGCCCCGGAGGGCCATGGACCCGAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCTGGAGGGCTTCGAAAGCGCGGGCATACCCCGGACCG 800
 L E T S W E A F E S A G I T P D A
 CGCGGGGCGACACCCGGCTGTTCATCGGCCGCTCTACGGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGACAGCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGCCTCCGGCCCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 30 35 35' 5' GTCACGGTCGACACCGCCTGCTCGTCGACTGGTCGCCCTGCACCAAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCTCGCTCGGCGAATGCTCGCTCGCCCTGGTCGGCGGTG 1050
 G Q S L R S G E C S L A L V G G
 TCACGGTGTGGCGTCGCCGGCGATTGTCGAGTTCTCCCGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGGGCGGGCGAAGGGCTTCGGCGCGGGCGGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTGGTCGAGGGCTCTCG 1200
 T S F A E G A G A L V V E R L S
 40 45 45' 5' ACGCGGAGCGCCACGGCCACACCGTCTCGCCCTCGTACCGGGCTCCCG 1250
 D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCTCGAACGGTCTGTCGGCGCCGAACGGCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCGTCATCCACCAAGGCCCTCGCGAACCGCAAACCTCACCCCG 1350
 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACCGCGTCGAGGGCGCACGGCACCCGCCCTCGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCGCAGGGCGCTCGCGACGTACGGACAGGACCGGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 50 55 55' 5' GCCCCCTGCTGCTCGCTCGTAAGTCGAAACATCGGGCACGCCAGGCC 1500
 P L L L G S L K S N I G H A Q A
 CCTCAGGGGTGCCCGGGATCATCAAGATGGTGCAGGCCATCGGCCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTCCACCGCGACGAGGCCGTCGCCGACGTCGACTG 1600
 E L P P T L H A D E P S P H V D W

GACGCCCGGTGCCCTCGAGCTCTGACGTGGCCGGCCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGGTGCCCGCGCCGCGCTGCCGTCTCGTCGTTGGCGTGGCGGCACG 1700
 T G F F R R A A V S S F S V S G T
 5 AACGCCACATCATCTTGAGGACGGACGGTCAGGACGGTCAGCGGCACG 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCAGTCAGGACGGACGGTCAGGACGGTCAGGAGCAGTCG 1800
 A G A I E A G P V E V G P V E A
 GACCGCTCCCCGGCGCCCGTCAAGCACCGGCGAGACTTCCGCTG 1850
 10 G P L P A A P P S A P G E D I P L
 CTCGTGTGCGCGCTCCCCGGAGGCACTCGACGAGCAGATCGGCCT 1900
 L V S A R S P E A L D E Q I G R L
 GCGCGCCATCTCGACACCGGGCCGGCGTCAGCGGGCGCCGTGGCGC 1950
 R A Y L D T G P G V D R A A V A
 15 AGACACTGGCCCGGTACGCACCTCACCCACCGGGCGTACTGCTCGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTCATCGGGCTCCCCCGCGGACCGAGCCGACGAACCTCGTCTT 2050
 D T V I G A P P A D Q A D E L V F
 CGTCTACTCCGGTCAGGGCACCCAGCATCCCGCAGGGCGAGCAACTCG 2100
 20 V Y S G Q G T Q H P A M G E Q L
 CGGCCGCGTTCCCCGTGTCGCCGATGCTGGCACACGCGCTCCGACGG 2150
 A A A F P V F A D A W H D A L R R
 CTCGACGACCCGACCCGACCGACCCCCACACGGAGCCAGCACACGCTCTT 2200
 L D D P D P H D P T R S Q H T L F
 25 CGCCCAACCAGGCGGGTTCACCGCCCTCCTGAGGTCTGGACATCACGC 2250
 A H Q A A F T A L L R S W D I T
 CGCACGCCGTACCGCCACTCGCTCGCGAGATCACCGCCCGTACGCC 2300
 P H A V I G H S L G E I T A A Y A
 GCGGGATCCTGCGCTCGACGACGCCGACCCCTGATCACCAACGCGTGC 2350
 30 A G I L S L D D A C T L I T T R A
 CCGCCTCATGCACACGCTCCGCCGCCATGGTCACCGTGTGA 2400
 R L M H T L P P P G A M V T V L
 CCAGCGAGGAGGAGGCCGTCAGGCGCTGCGGCCGGCGTGGAGATGCC 2450
 T S E E E A R Q A L R P G V E I A
 35 GCGGTCTCGGCCGACTCCGTCGTGCTCTCGGGCGACGAGGACGCCGT 2500
 A V F G P H S V V L S G D E D A V
 GCTCGACGTCGACAGCGGCTCGGCATCCACCACCGTCTGCCCGCGCCGC 2550
 L D V A Q R L G I H H R L P A P
 ACGCGGGCCACTCCGCGCACATGGAACCCGTGGCCCGAGCTGCTGCC 2600
 40 H A G H S A H M E P V A A E L L A
 ACCACTCCGAGCTCCGTTACGACCGGCCACACGCCATCCGAACGA 2650
 T T R E L R Y D R P H T A I P N D
 CCCCACCAACCGCCGAGACTGGCCGAGCAGGTCTGGCAACCCCGTGTGT 2700
 P T T A E Y W A E Q V R N P V L
 45 TCCACGCCACACCCAGCGGTACCCCGACGCCGTCTCGTGGAGATGCC 2750
 F H A H T Q R Y P D A V F V E I G
 CCCGGCCAGGACCTCTCACCGCTGGTCAGGGCATGCCCTGAGAACGG 2800
 P G Q D L S P L V D G I A L Q N G
 CACGGCGGACGAGGTGCACCGCGCTGCACACCGCGCTGCCCGCTTCA 2850
 50 T A D E V H A L H T A L A R L F
 CACCGCGGCCACGCTCGACTGGTCCCGCATCCTCGCGGTGCTCGCGG 2900
 T R G A T L D W S R I L G G A S R
 CACGACCCCTGACGTCCCTCGTACCGCTCCAGCGCGCTCCACTGGAT 2950
 H D P D V P S Y A F Q P R P Y W I
 55 CGAGTCGGCTCCCCCGGCCACGGCGACTCGGGCCACCCCGTCTCGCA 3000
 E S A P P A T A D S G H P V L G
 CCGGAGTCGCCGTCCCCGGGTGCGCGGGCGGGTGTACGGGTCCCCGTG 3050
 T G V A V A G S P G R V F T G P V
 CCCCCCGGTGCGGACCGCCGGTGTTCATCGCCGAACCTGGCGCTCGCCGC 3100
 60 P A G A D R A V F I A E L A L A A

CGCCGACGCCACCGACTGCCACGGCGAACAGCTGACGTACCTCCG 3150
 A D A T D C A T V E Q L D V T S
 TGCCCGCGGATCCGCCCGCGCAGGGCCACCGCGCAGACCTGGTCGAT 3200
 P G G S A R G R A T A D T W V D
 5 SAAACCGCCGCCGACGGCGCCCTCACCGTCCACACCCCGTCGG 3250
 E P A A D G R R F T V H T R Y G
 CGACGCCCGTGGACGCTGCACGCCGAGGGGTTCTCCGCCCGCGCG 3300
 D A P W T L H A E G V L R P G R
 TCCCCAGCCGAAGCCGTCACCCGCTGGCCCCCGCCGGCGCGGTG 3350
 V P Q P E A V D T A W P P P G A V
 CCCGCAGGGCTGCCCGGGCGTGGCAGCGCGGACCAAGGTCTCGT 3400
 P A D G L P G A W R R A D Q V F V
 10 SGAAGCCGAAGTCGACAGCCCTGACGGCTCGTGGCACACCCGACCTGC 3450
 E A E V D S P D G F V A H P D L
 TCGACCGGGTCTTCTCCGCGGTGACGGGAGGCCAGCCGACCGGA 3500
 L D A V F S A V G D G S R Q P T G
 TGGCGCACCTCGCGGTGACCGCTGGACGCCACCGTGCTGCGCGCTG 3550
 W R D L A V H A S D A T V L R A C
 15 CCTCACCCGCCGACAGTGGTGTGAGCTGCCGCCTCGACGGTG 3600
 L T R R D S G V V E L A A F D G
 CCGGAATGCCGGTGTCAACCGCGGAGTCGGTGACGCTGGCGAGGTGCG 3650
 A G M P V L T A E S V T L G E V A
 TCGGCAGGGGATCCGACGAGTCGGACGGTCTGCTCGGTTGAGTGGTT 3700
 20 S A G G S D E S D G L L R L E W L
 SCCCCGGCGGAGGCCACTACGACGGTGCCGACGAGCTGCCGAGGGCT 3750
 P V A E A H Y D G A D E L P E G
 ACACCCCTCATCACCGCCACACACCCGACGACCCGACGACCCACCAAC 3800
 Y T L I T A T H P D D P D D P T N
 25 CCCCCACAACACACCCACACGGCACCCACACACAAACACCGCTCTCAC 3850
 P H N T P T R T H T Q T T R V L T
 CGCCCTCCAACACCCACTCATCACCAACACACCCCTCATCGTCCACA 3900
 A L Q H H L I T T N H T L I V H
 30 CCACCAACGGCCCCCAGGGCGCCGTCACCGGCTCACCCGACCGCA 3950
 T T T D P P G A A V T G L T R T A
 CAAAACGAACACCCGGCGATCCACCTCATCGAAACCCACACCCACCA 4000
 Q N E H P G R I H L I E T H H P H
 CACCCACTCCCCCTACCCAACTCACCAACCTCACCAACCCACCTAC 4050
 35 T P L P L T Q L T T L H Q P H L
 GCCTCACCAACACCCCTCCACACCCCCCCTCACCCCATCACCAAC 4100
 R L T N N T L H T P H L T P I T T
 CACCAACACCCACCAACCCACCCCCAACACCCCCACCCCTCAACCCCAA 4150
 H H N T T T T P N T P P L N P N
 CCACGCCATCCTCATCACGGCGGCTCCGGCACCTCGCCGGCATCTCG 4200
 H A I L I T G G S G T L A G I L
 40 CCCGCCACCTCAACCAACCCCCACACCTACCTCTCTCCGCACACCA 4250
 A R H L N H P H T Y L L S R T P P
 CCCCCCACCACACCCGGCACCCACATCCCTGCGACCTCACCGACCCAC 4300
 P P T T P G T H I P C D L T D P T
 CCAAATCACCAAGCCCTCACCCACATACCACAACCCCTCACCGGATCT 4350
 45 Q I T Q A L T H I P Q P L T G I
 TCCACACCGCCGCCACCTCGACGACGCCACCCCTACCAACCTACCCCC 4400
 F H T A A T L D D A T L T N L T P
 CAACACCTCACCAACCCCTCAACCCAAAGCCGACGCCGCTGGCACCT 4450
 Q H L T T T L Q P K A D A A W H L
 50 CCACCAACACCCAAAACCAACCCCTCACCCACTTCGTCTACTCCA 4500
 H H H T Q N Q P L T H F V L Y S
 GCGCCGCCGCCACCCCTCGGCAGCCCCGGCCAAGCCAACCTACGCCGCC 4550
 S A A A T L G S P G Q A N Y A A A
 AACGCCCTCCTCGACGCCCTCGCCACCCACCCAAGGACAACC 4600
 55 N A F L D A L A T H R H T Q G Q P

CGCCACCAACCACATCGCCTGGGGCATGTGGCACACACCACCAACTCACCA 4650
 A T T I A W G M W H T T T T L T
 GCCAACTCACCGACAGCGACCGCGACCCGATCCGCCGCGGGCTTCCTG 4700
 S Q I T C S D R D R I R R G G F I
 5 CGATCTCGGACGACGAGGGCATGC
 P I S D D E S M

The *AvrII-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 12 of rapamycin is shown below.

10 SCATGCGGCTGTACGAGGCGGCACGGCGCACCGGAAGTCCCCTGGTGGTG 50
 M R L Y E A A R R T G S P V V V
 GCGGCCGCGCTCGACGACGCGCCGGACGTGCGCTGCTGCGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCGTGCCTGGGACCGCTCTCGCCGAC 150
 15 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCGCCGACGCCCTCGCGTTG 200
 R S P C C P T T S A P T P P S R S
 TCCTGGAACAGCACCGCCACCGTGCTCGGCCACCTGGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 20 CCCGGCGACGACGACGTTCAAGGAACCTCGGCATCGACTCGCTACCGCGG 300
 P A T T T F K E L G I D S L T A
 TCCAGCTGCGCAACCGCGCTGACCAACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACCCCGCGCGCTCGCCCGAGACTCGG 400
 25 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCGGTACCCCGCGCCCGTGCACGGCCGGACCGCGGCCA 450
 D E L A G T R A P V A A R T A A
 CCGCGGCCGCCACGACGAACCGCTGGCGATCGTGGCATGGCTGCGT 500
 T A A A H D E P L A I V G M A C R
 30 CTGCCGGGGGGGTCCGTCGCCACAGGAGCTGTGGCGTCTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATACGGAGTTCCCCCGGGACCGCGGCTGGACGTGG 600
 G T D A I T E F P A D R G W D V
 35 ACAGCGCTCTACGACCCGGACCCCGACGCGATCGGCAAGACCTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCCTCGACGGTGGGACCGGCTTCGACCGGGCTTCCGG 700
 H G G F L D G A T G F D A A F F G
 GATCAGCCCCCGCGAGGCCCTGGCCATGGACCCGCAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 40 TGGAGACGTCCCTGGGAGGCCTCGAAAGCGCGGGCATCACCCGGACGCG 800
 L E T S W E A F E S A G I T P D A
 GCGCGGGGCAGCGACACCGCGTGTTCATCGCGCGTTCTCCTACGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGCGACAGGGTGCAGACCA 900
 45 G T G A D T N G F G A T G S Q T
 CGCTGCTCTCCGGCCGCCCTCGTACTTCTACGGTCTGGAGGGCCCTTCG 950
 S V L S G R L S Y F Y G L E G P S
 GTCACGGTCGACACCGCCCTCGTCGTCACTGGTCCCCCTGCACCCAGGC 1000
 50 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTCGCGCTCGGGCGAATGCTCGCTCGCCCTGGTCGGCGGTG 1050
 G Q S L R S G E C S L A L V G G
 TCACGGTGTGGCGCTCGCCGGGGATTCTCGCGAGTTCTCCGGACCGCG 1100
 V T V M A S P G G F V E F S R Q R
 55 G G G C T C G C G C G A C G G G C G G G A A G G C G T T C G G C G G G C G G A C G G 1150
 G L A P D G R A K A F G A G A D G
 TACGGAGCTTCCCGAGGGCGCCGGTGCCTGGTGGTGGACGGCTCTCCG 1200
 T S F A E G A G A L V V E R L S
 ACGCGGAGCGCCACCGCCACACCGTCCCTCGCCCTCGTACGCGGCTCCGCG 1250
 D A E R H G H T V L A L V R G S A

GCTAACTCCGACGGCGCGTCGAACGGTCTGTCGGCGCGAACGGCCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 CCAGGAACCGTCACTCCACCAGGCCCTCGCGAACCGAAACTCACCCCCG 1350
 Q E R V I H Q A L A N A K I T P
 5 CCGATGTCGACCGCGTCGAGGCCACGGCACCCGCGCTCGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGAGGCCAGGCCCTCGCGACGTACGGACAGGACCGGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 GCCCCCTGCTGCTCGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCCG 1500
 10 P L L L G S L K S N I G H A Q A
 CGTCAGGGGGTCCGCCGGGATCATCAAGATGGTGCAGGCCATCCGGCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCGACACTGCACGCCGAGGCCCTGCCGCACGTGACTG 1600
 E L P P T L H A D E P S P H V D W
 15 GACGGCCGGTCCGCTCGAGCTCTGACGTCGGCCCGCCGTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 CCGCTGCCCTAGGCCGGCAGGCCGTGTCGTCCTCGGGATCAGTGGCACC 1700
 T G R P R A G V S S F G I S G T
 AACGCCACGTCACTGGAAAGCGCACCCCCCACTCAGCCTGCCGACAA 1750
 20 N A H V I L E S A P P T Q P A D N
 CGCGGTGATCGAGCCGGCACCGGAGTGGGTGCCGTTGGTGATTCCGGCCA 1800
 A V I E R A P E W V P L V I S A
 GGACCCAGTCGGCTTGACTGAGCACGAGGGCCGGTTCGTCGATCTG 1850
 R T Q S A L T E H E G R L R A Y L
 25 GCGCGTCGCCCGGGGTGGATATGCCGGCTGTGGCATCGACGCTGGCGAT 1900
 A A S P G V D M R A V A S T L A M
 GACACGGTCGGTGGTCAGCACCGTGCTGCTGGAGATGACACCG 1950
 T R S V F E H R A V L L G D D T
 30 TCACCGGCACCGCTGTCGACCCCTCGGGCGGTGTTGCTCTCCGGGA 2000
 V T G T A V S D P R A V F V F P G
 CAGGGGTCCAGCGTGCCTGGCATGGTGAGGAACCTGGCCGCCGTTCCC 2050
 Q G S Q R A G M G E E L A A A F P
 CGTCTTCGCGCGGATCCATCAGCAGGTGTGGGACCTGCTCGATGTGCCG 2100
 V F A R I H Q Q V W D L L D V P
 35 ATCTGGAGGTGAACGAGACCGTTACGCCAGCCGGCCCTGTCGAATG 2150
 D L E V N E T G Y A Q P A L F A M
 CAGGTGGCTCTGTCGGGCTGCTGGAAATCGTGGGTGACGCCGGACGC 2200
 Q V A L F G L L E S W G V R P D A
 GGTGATCGGCCATTGGTGGGTGAGCTTGGCGCTGCGTATGTGTCGGGG 2250
 40 V I G H S V G E L A A A Y V S G
 TGTGGTCGTTGGAGGATGCCCTGCACTTGGTGTGGCGCGGGCTCGT 2300
 V W S L E D A C T L V S A R A R L
 ATGCAGGCTCTGCCGCCGGTGGGTGATGGTCGCTGTCGGGCTCGGA 2350
 M Q A L P A G G V M V A V P V S E
 45 GGATGAGGCCGGCCGTGCTGGGTGAGGTGTGGAGATGCCGCCGTCA 2400
 D E A R A V L G E G V E I A A V
 ACGGCCCGTCGTCGGTGGTCTCTCCGGTGTGGAGGCCGCCGTGCTGCAG 2450
 N G P S S V V L S G D E A A V L Q
 50 GCCGCCGGAGGGGCTGGGAAGTGGACGCCGTGGCGACCAGCCAGCGTT 2500
 A A E G L G K W T R L A T S H A F
 CCATTCCGCCGTATGGAACCCATGCTGGAGGAGTTCCGGCGGTGCGCCG 2550
 H S A R M E P M L E E F R A V A
 AAGGCCCTGACCTACCGGACCCCGCAGGTCTCCATGCCGTTGGTGTACAG 2600
 55 E G L T Y R T P Q V S M A V G D Q
 ATGACCACCGCTGAGTACTGGTGCAGGCCGAGGTCCGGACACGGTCCGGT 2650
 V T T A E Y W V R Q V R D T V R F
 CGGCGAGCAGGTGGCCTCGTACGAGGACGCCGTGTCGAGCTGGGTG 2700
 G E Q V A S T E D A V F V E I G
 60 CCGACCGGTCACTGCCGCCCTGGTCGACGGTGTGCGGATGCTGCCACGGC 2750
 A D R S L A R L V D G V A M L H G

GACCACGAAATCCAGGCCCGATCGGCGCCCTGGCCACCTGTATGTCAA 2800
 D H E I Q A A I G A L A H L Y V N
 CGCGCTCACGGTCGACTGGCCCGCGCTCCTGGCGATGCTCCGGCAACAC 2850
 G V T V D W P A L L G D A P A T
 5 GGGTGTGACTGGACCTTCGACATACGCCCTCAGCACCGCGCTACTGGCTC 2900
 R V I D L P T Y A F Q H Q R Y W L
 GAGTCGGCTCCCCCGGCCACGGCCACTCGGGCACCCCGTCCTCGGCAC 2950
 E S A P P A T A D S G H P V L G T
 CGGACTCCCCGTGGCCGGTCGCCGGGGTCTTACGGGTCCCCTGC 3000
 10 G V A V A G S P G R V F T G P V
 CGCGCGGTGCGGACCGCGCGGTGTTCATCGCCGAAGTGGCGCTGCCGCC 3050
 P A G A D R A V F I A E L A L A A
 GCCGACGCCACCGACTCCGCCACGGTCGAACAGCTCGACGTACCTCCGT 3100
 15 A D A T D C A T V E Q L D V T S V
 GCCCGGGCGATCCGCCCGCGGCAAGGGCACCGCGCAGACCTGGTCGATG 3150
 P G G S A R G R A T A Q T W V D
 AACCCCGCCGCCGACGGCGCGCGCTTACCGTCCACACCCCGCGTCGGC 3200
 E P A A D G R R F T V H T R V G
 GACGCCCGTGGACGCTGCACGCCGAGGGGGTTCTCCGCCCGGCCCGT 3250
 20 D A P W T L H A E G V L R P G R V
 GCCCCAGCCCGAAGCCGTCGACACCGCTGGCCACACCCCGCGGTGC 3300
 P Q P E A V D T A W P P P G A V
 CCGCGGACGGGCTGCCGGGCGTGGCGACGCCGCGGACAGGTCTCGTC 3350
 P A D G L P G A W R R A D Q V F V
 25 GAAGCCGAACTGACAGCCCTGACGGCTTGTGGCACACCCGACCTGCT 3400
 E A E V D S P D G F V A H P D L L
 CGACCGGGTCTCTCCGCCGGTCGGCGACGGGAGCCGCCAGCCGACCGGAT 3450
 D A V F S A V G D G S R Q P T G
 GGCGCGACCTCGCGGTGCACGCCGCGGACGCCACCGTGTGCGCGCCTGC 3500
 30 W R D L A V H A S D A T V L R A C
 CTCACCCGCCGCGACAGTGGTGTGGAGCTGCCGCCCTCGACGGTGC 3550
 L T R R D S G V V E L A A F D G A
 CGGAATGCCGGTGCTCACCGCGGAGTCGGTACGCTGGCGAGGTGCGT 3600
 G M P V L T A E S V T L G E V A
 35 CGGCAGGCAGGATCCGACGAGTCGGACGGTCTGCTCGGCTTGAGTGGTTG 3650
 S A G G S D E S D G L L R L E W L
 CCGGTGGCGAGGCCACTACGACGGTGCCGACGAGCTGCCGAGGGCTA 3700
 P V A E A H Y D G A D E L P E G Y
 CACCCCTCATACCGCCACACACCCGACGACCCGACGACCCACCAACC 3750
 40 T L I T A T H P D D P D D P T N
 CCCACAACACACCCACACGACCCACACACAAACACCGTCTCCTCACC 3800
 P H N T P T R T H T Q T T R V L T
 GCCCTCCAACACCCACCTCATCACCAACACACCCCTCATCGTCCACAC 3850
 A L Q H H L I T T N H T L I V H T
 45 CACCAACGACCCCCCAGGGCGCCGCCGTACCCGGCTCACCGCACCGAC 3900
 T T D P P G A A V T G L T R T A
 AAAACGAAACACCCGGCCATCCACCTCATCGAAACCCACACCCAC 3950
 Q N E H P G R I H L I E T H H P H
 ACCCCACTCCCCCTCACCCAACTCACCCACCCCTCACCAACCCACCTACG 4000
 50 T P L P L T Q L T T L H Q P H L R
 CCTCACCAACACACCCCTCACACCCCCCACCTCACCCCCCATCACCAAC 4050
 E T N N T L H T P H L T P I T T
 ACCACAAACACCCACCAACCCACACCCCCAACACCCACCCCTCAACCCCAAC 4100
 H H N T T T T P N T P P L N P N
 55 CACGCCATCCATCACCGGGGGTCCGGCACCCCTGCCGGCATCCTCGC 4150
 H A I L I T G G S G T L A G I L A
 CGGCCACCTAACCCACACCCACACCTACCCCTCTCCCGCACACCCAC 4200
 R H L N H P H T Y L L S R T F P
 CCCCCACCAACACCCGGCACCCACATCCCTGCGACCTCACCGACCCAC 4250
 60 P P T T P G T H I P C D L T D P T

CAAATCACCAAGCCCTCACCCACATACCAACCAACCCCTCACCGGCATCTT 4300
 Q I T Q A L T H I P Q P L T G I F
 CCACACCGCCGCCACCCCTCGACGACGCCACCCCTCACCAACCTCACCCCCC 4350
 H T A A T L D D A T L T N L T P
 5 AACACCTCACCAACCCCTCAACCCAAAGCCGACGCCGCCTGGCACCTC 4400
 Q H L T T L Q P K A D A A W H L
 CACCAACACCCAAAACCAACCCCTCACCCACTTCGTCCTACTCCAG 4450
 H H H T Q N Q P L T H F V L Y S S
 CGCCGCCGCCACCCCTCGGCAGCCCCGGCCAAGCCAACCTACGCCGCC 4500
 10 A A A T L G S P G Q A N Y A A A
 ACGCCTTCCTCGACGCCCTCGCCACCCACCCGACACCCAAAGGACAACCC 4550
 N A F L D A L A T H R H T Q G Q P
 GCACACCACCATCGCCTGGGCATGTGGCACACCACCACTCACCAG 4600
 A T T I A W G M W H T T T L T S
 15 CCAACTCACCGACAGCGACCGCAGCCATCCGCCGCCGGCTTCCTGC 4650
 Q L T D S D R D R I R R G G F L
 CGATCTCGGACGACGAGGGCATGC
 P I S D D E G M

20 The *AvrII-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module
 13 of rapamycin is shown below.

GCATGGGCTGTACGAGGGCGGCACGGCGCACCGGAAGTCCGTGGTGGT 50
 M R L Y E A A R R T G S P V V V
 25 GCGGCCGCCCTCGACGACGCCGGACGTGCCGCTGCTGCGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCTGCCGCCGTCCGGGAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCGCCACGCCCTCCCTCGCGTCG 200
 R S P C C P T T S A P T P P S R S
 30 TCCCTGAAACAGCACCGCCACCGTGTCTGGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCAGCACGACGTTCAAGGAACCTCGGCATCGACTCGCTACCGCGG 300
 P A T T T F K E L G I D S L T A
 TCCAGCTGCGCAACCGCCTGACCAACGGGACCGGGTACGCCAACGCC 350
 35 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTTCGACTTCCGACGCCGCCGCGCTCGCCGCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGCCCGGTACCCGCGCCCGTCCGGCCCGGACCGCGGCCA 450
 D E L A G T R A P V A A R T A A
 40 CCGCGGCCGCCACGACGAACCGCTGGCGATCGTGGCATGGCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGGGGGGTCGCGTCGCCAACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATCACGGAGTCCCCGGGACCGGGCTGGACGTGG 600
 45 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCCGACCGCGATCGGCAAGACCTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCTCGACGGTGGCACCGGCTTCGACGCCGGTCTTCGG 700
 H G G F L D G A T G F D A A F F G
 50 GATCAGCCCGCGGAGGGCCATGGACCCGACGCCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCTGGGAGGGTTCGAAAGCGCGGGCATACCCCGGACCGCG 800
 L E T S W E A F E S A G I T P D A
 GCGCGGGGCGACACCCCGTGTTCATCGGCCGTTCTCCTACGGGTA 850
 55 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGGTTGGGCCGACAGGGTCGGACAGACCA 900
 G T G A D T N G F G A T G S Q T
 CGGTGCTCTCCGGCCGCTCTCGTACTTCTACGGTCTGGAGGGCCCTTCG 950
 S V L S G R L S Y F Y G L E G P S

5' T C A C G G T C G A C A C C C C T G C T C G T C A C T G G T C G C C C T G C A C C A G G C 1000
 V T V D T A C S S S L V A L H Q A
 A G G G C A G T C C C T G C G C T C G G G C A A T G C T C G C T C G C C C T G G T C G G C G G T G 1050
 S Q S L R S G E C S L A L V G G
 5 T C A C G G T G A T G G C T C G C C C G G C G G A T T C G T C G A G G T T C T C C C G G C A G C G C 1100
 V T V M A S P G G F V E F S R Q R
 G G G C T C G C C C G G A C G G G C G G A A G G C G T C G G C G C G G C G C G G A C G G 1150
 G L A P D G R A K A F G A G A D G
 T A C G A G C T T C G C C G A G G G C G C C G G T G C C C T G G T G G T C G A G C G G C T C C G 1200
 10 T S F A E G A G A L V V E R L S
 A C G C G G A G G C C A C G G C C A C C C G T C C T C G C C C T C G T A C G C G C T C C G C G 1250
 D A E R H G H T V L A L V R G S A
 G C T A A C T C C G A C G G C C G G T C G A C G G T C T G T C G G C G C C G A C G G C C C C T C 1300
 A N S D G A S N G L S A P N G P S
 15 C C A G G A A C G C G T C A T C C A C C A G G C C C T C G C G A A C G C G A A A C T C A C C C C C G 1350
 Q E R V I H Q A L A N A K L T P
 C C G A T G T C G A C G C G G T C G A G G G C C A C G G C A C C G G C A C C C G C C T C G G C G A C 1400
 A D V D A V E A H G T G T R L G D
 20 C C C A T C G A G G G C C A G G C G C T G C T C G C G A C G T A C G G A C A G G A C C G G G C G A C 1450
 P I E A Q A L L A T Y G Q D R A T
 G C C C C T G C T G C T C G G C T C G C T G A A G T C G A A C A T C G G G C A C G C C C A G G C C G 1500
 P L L L G S L K S N I G H A Q A
 C G T C A G G G G T C G C C G G A T C A A G A T G G T G C A G G G C A T C C G G C A C G G G 1550
 A S G V A G I I K M V Q A I R H G
 25 G A A C T G C C G C C G A C A C T G C A C G C G G A C G A G G C C G T C G C G C A C G T C G A C T G 1600
 E L P P T L H A D E P S P H V D W
 G A C G G C C G G T G C C G T C G A G G C T C T G A C G T C G G C C C G G C G T G G C C G G G G A 1650
 T A G A V E L L T S A R P W P G
 C C G G T C G C C C T A G G C G G G C G G G C G T G C G T C C T C G G A G T C A G C G G C A C C 1700
 30 T G R P R R A G V S S F G V S G T
 A A C G C C C A C G T C A T C C T G G A G A G C G C A C C C C C C G T C A G C C C G C G G A G G A 1750
 N A H V I L E S A P P A Q P A E E
 G G C G C A G C C T G T T G A G A C G C C G G T G G T G G C C T C G G A T G T G C T G C C G C T G G 1800
 A Q P V E T P V V A S D V L P L
 35 P G A T A T C G G C C A A G A C C C A G C C C G C C T G A C C G A A C A C G A A G A C C G G C T G 1850
 V I S A K T Q P A L T E H E D R L
 C G C G C C T A C C T G G C G G C T C G C C C G G G C G G A T A T A C G G G C T G T G G C A T C 1900
 R A Y L A A S P G A D I R A V A S
 G A C C C T G G C G G T G A C A C G G T C G G T G T T C G A G C A C C G C G C C G T A C T C C T G 1950
 40 T L A V T R S V F E H R A V L L
 G A G A T G A C A C C G T C A C C G G C A C C C G G G T G A C C G A C C C C A G G A T C G T G T T 2000
 G D D T V T G T A V T D P R I V F
 G T C T T C C C G G G C A G G G G T G G C A G T G G C T G G G G A T G G G C A G T G C A C T G C G 2050
 V F P G Q G W Q W L G M G S A L R
 45 C G A T T C G T C G G T G G T T C G C C G A G C G G A T G G C G A G T G T G C G C G C G C G T 2100
 D S S V V F A E R M A E C A A A
 T G C G C G A G T T C G T G G A C T G G G A T C T G T T C A C G G T T C T G G A T G A T C C G G C G 2150
 L R E F V D W D L F T V L D D P A
 G T G G T G G A C C G G G T T G A T G T G G T C C A G C C C G C T T C C T G G G C G A T G A T G G T 2200
 50 V V D R V D V V Q P A S W A M M V
 T T C C C T G G C C C C G G T G G C A G G C G G C C G G T G C G G G C G G A T G C G G T G A 2250
 S L A A V W Q A A G V R P D A V
 T C G G C C A T T C G C A G G G T G A G A T C G C C G C A G C T T C T G T G G C G G G T G C G G T G 2300
 I G H S Q G E I A A A C V A G A V
 55 T C A C T A C G C G A T G C C C C C G G A T C G T G A C C T T C G C G C A G C C A G G C G A T C G C 2350
 S L R D A A R I V T I R S Q A I A
 C C G G G G C C T G G C G G C C C G G G C G G A T G G C A T C C G T C G C C C T G C C C G C G C 2400
 P Q L A G R G A M A S V A L P A
 60 A G G A T G T C G A G C T G G T C G A C G G G G C T G G A T C G C C G C C C A C A A C G G G C C C 2450
 C D V E L V D G A W I A A H N G P

GCCTCCACCGTCACTCGCGGGCACCCCGGAAGCGGTCAACCATCTCTCAC 2500
 A S T V I A G T P E A V D H V L T
 CGCTCATGAGGCACAAAGGGTGCAGGGTCCGGCGGATCACCGTCACTATG 2550
 A H E A P G V R V R I T V D Y
 5 CCTCGCACACCCCGCACGTGAGCTGATCCCGACGAACTACTCGACATC 2600
 A S H T P H V E L I R D E L L D I
 ACTAGCGACAGCAGCAGCAGACCCCGCTCGTGCCTGGCTGTCGACCGT 2650
 T S D S S S Q T P L V P W L S T V
 GGACGGCACCTGGTCACTGGGAGTACTGGTACCGGA 2700
 10 D G T W V D S P L D G E Y W Y R
 ACCTGCGTGAACCGGTCGGTTCCACCCCGCTCAGCCAGTTGCAGGCC 2750
 N L R E P V G F H P A V S Q L Q A
 CAGGGCGCACCCGTGTCGAGGTCAAGCCAGCCGGTGTGCA 2800
 Q G D T V F V E V S A S P V L L Q
 15 GGCAGATGGACGACGATGTCGTACCGTTGCCACGCTCGTCGTGACGACG 2850
 A M D D D V V T V A T L R R D D
 GCGACGCCACCCCGATGCTCACCGCCCTGGCACAGGCCTATGTCACGGC 2900
 G D A T R M L T A L A Q A Y V H G
 GTCACCGTCGACTGGCCCGCCATCCTCGGACCACCAACCCGGTACT 2950
 20 V T V D W P A I L G T T T T R V L
 GGACCTTCCGACCTACGCCCTCCAACACCAGCGGTACTGGCTCGAGTCGG 3000
 D L P T Y A F Q H Q R Y W L E S
 CTCCCCCGGCCACGGCGACTCGGGCCACCCCGTCCCTGGCACCGGAGTC 3050
 A P P A T A D S G H P V L G T G V
 25 GCGTCGCGGGTCCGGGGCCGGGTGTTACCGGTCCCCGTGCCCGCCGG 3100
 A V A G S P G R V F T G P V P A G
 TGCGGACCGCGCGGTGTTACCGCGACTGGCGCTGCCGCCGGCGACG 3150
 A D R A V F I A E L A L A A D
 CCACCGACTGCCACGGTCAACAGCTCGACGTACCTCCGTGCCCGCC 3200
 30 A T D C A T V E Q L D V T S V P G
 GGATCCGCCCGGGCAGGGCCACCGCGCAGACCTGGTCGATGAACCCGC 3250
 G S A R G R A T A Q T W V D E P A
 CGCCGACGGCGGCCGCTTACCGTCCACACCCGGTCCGGCACGCC 3300
 A D G R R R F T V H T R V G D A
 35 CGTGGACGCTGCACGCCGAGGGGGTTCTCCGCCCGCGTGGCCAG 3350
 P W T L H A E G V L R P G R V P Q
 CCCGAAGCGTCGACACCGCCTGGCCCCCGCCGGCGCGGTGCCCGGGA 3400
 P E A V D T A W P P P G A V P A D
 CGGGCTGCCGGGGCGTGGCGACCGCGCGGACCGAGGTCTCGTCAAGCCG 3450
 40 G L P G A W R R A D Q V F V E A
 AAGTCGACAGCCCTGACGGCTTCTGGCACACCCCGACCTGCTCGACGCG 3500
 E V D S P D G F V A H P D L L D A
 GTCCTCTCCGGTGGCGACGGGAGCCGCGACGGATGGCGCGA 3550
 V F S A V G D G S R Q P T G W R D
 45 CCTCGCGGTGACCGCGTCCGGACGCCACCGTGTGCGGCCCTGCCACCC 3600
 L A V H A S D A T V L R A C L T
 GCCCGCACAGTGGTCTGGAGCTCGCCGCTTCGACGGTGGCGGAATG 3650
 R R D S G V V E L A A F D G A G M
 CCGGTGCTCACCGCGGAGTCGGTACGCTGGCGAGGTGCGTCCGGCAGG 3700
 50 P V L T A E S V T L G E V A S A G
 CGGATCCGACAGTCGGACGGTCTGCTTGGCTGAGTGGTTGCCGGTGG 3750
 G S D E S D G L L R L E W L P V
 CGGAGGCCACTACGACGGTGCACGAGCTGCCGAGGGCTACACCCCTC 3800
 A E A H Y D G A D E L P E G Y T L
 55 ATCACCGCCACACACCCCGACGACCCCGACGACCCCCACCAACCCCAA 3850
 I T A T H P D D P D D P T N P H N
 CGACACCCGACCGCACACACACACACACACACACACACACACAC 3900
 T P T F T H T Q T T R V L T A L
 AACACCACTCATCACCAACCAACACACCCCATCGTCCACACCAACACC 3950
 60 Q H H L I T T N H T L I V H T T T

GACCCCCCAGGCGCCGCCGTACCGGCCTCACCGCACCGAACGA 4000
 D P P G A A V T G L T R T A Q N E
 ACACCCCCGGCCGATCCACCTCATCGAAACCCACCACCCCCACACCCCCAC 4050
 H P G R I H L I E T H H P H T P
 5 TCCCCCTCACCCAACTCACCCACCCCTCCACCAACCCACCTACGCCSACC 4100
 L P L T Q L T T L H Q P H L R L T
 AACAAACACCCCTCCACACCCCCACCTCACCCCCATCACCCACCAACAA 4150
 N N T L H T P H L T P I T T H H N
 CACCAACCAACCAACCCCCAACACCCCCACCCCTCAACCCAAACCAACGCCA 4200
 10 T T T T T P N T P P L N P N H A
 TCCTCATCACCGGGGCGCTCCGGCACCCCTCGCCGGATCCCTGCCGGCAC 4250
 I I I T G G S G T L A G I L A R H
 CTCAACCCACCCCCACACCTACCTCCCTCCCGCACACCCACCCCCCAC 4300
 15 L N H P H T Y L L S R T P P P P T
 CACACCCGGCACCCACATCCCCCTCGGACCTCACCGACCCACCCAAATCA 4350
 T P G T H I P C D L T D P T Q I
 CCCAAGCCCTCACCCACATACCAACCCCTCACCGCATTTCCACACC 4400
 T Q A L T H I P Q P L T G I F H T
 GCGGCCACCCCTCGACGACGCCACCCCTACCAACCTCACCCCCAACACCT 4450
 20 A A T L D D A T L T N L T P Q H L
 CACCAACCCCTCCAACCCAAAGCCGACGCCGCTGGCACCTCCACCCACC 4500
 T T T L Q P K A D A A A W H L H H
 ACACCCAAAACCAACCCCTACCCACTTCGTCCTCTACTCCAGCGCCGCC 4550
 H T Q N Q P L T H F V L Y S S A A
 25 GCCACCCCTCGGCAGCCCCGGCCAAGCCAACCTACGCCGCCAACGCCCT 4600
 A T L G S P G Q A N Y A A A A N A F
 CCTCGACGCCCTCGCCACCCACCGCCACACCCAGGACAACCCGCCACCA 4600
 L D A L A T H R H T Q G Q P A T
 CCATGCCCTGGGCATGTGGCACACCACCAACTCACCAAGCCAACTC 4700
 30 T I A W G M W H T T T T L T S Q L
 ACCGACAGCGACCGCGACCGCATCCGCCGCGGCTTCCTGCCGATCTC 4750
 T D S D R D R I R R G G F L P I S
 GGACGACGAGGGCATGC
 D D E G M
 35

The *NheI-Xhol* hybrid FK-506 PKS module 8 containing the AT domain of module 12 of rapamycin is shown below.

GCATGGGCTGTACGAGGGCGCACGGCGACCGGAAGTCCCCTGGTGGTGGT 50
 M R L Y E A A A R R T G S P V V V
 40 CGGGCCGCGCTCGACGACGCCGGACGTGCCGCTGCTGCCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 GCGTACGACCGTCCGGCGTCCGGCGTCCGGGAACGCTCTCGCCGACC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCCGACGCCCTCCCTCGCGTTCG 200
 45 R S P C C P T T S A P T P P S R S
 TCCCTGAAACAGCACCGCCACCGTCTGGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACCTCGGATCGACTCGCTCACCGCGG 300
 P A T T T F K E L G I D S L T A
 50 TCCAGCTGCCAACCGCGCTGACCAACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTCGACTTCCGACGCCGCGCGCTGCCGCGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCGGTACCCGCGCCCGTGCAGGCCCGGACCGCGGCCA 450
 55 D E L A G T R A P V A A R T A A
 CGCGGGCCGCGCACGAAACCGCTGGCGATCGTGGCATGGCGTGGCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCCGGCGGGGTGGCTGCCACAGGAGCTGTGGCGTCTCGTCGCGTC 550
 L P G G V A S P Q E L W R L V A S

CGGCACCGACGCCATCACGGAGTTCCCCGGACCGCGGCTGGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACGCGCTCTACGACCCGGACCCCGACCGATCGGCAAGACCTTCGTCCGG 650
 D A L Y D P D F D A I G K T F V R
 5 CACGGCGGCTTCCTCGACGGTGCACGGCTTCGACGCCGCTTCGG 700
 H G G F L D G A T G F D A A A F F G
 GATCAGCCCCCGCGAGGCCCTGCCATGGACCCGCAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCTGGAGGCCTCGAAAGCCGGCATCACCCGGACGCG 800
 10 L E T S W E A F E S A G I T P D A
 GCGCGGGGACGCCACCCGGCTGTTCATCGGCGCTTCTCTACGGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGCAGACCA 900
 G T G A D T N G F G A T G S Q T
 15 GCGTGCTCTCCGGCCGCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 S V L S G R L S Y F Y G L E G P S
 GTACCGGTGACACCGGCTGCTCGTCGACTGGTCGCCCTGCACCAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCCTGCGCTCGGGCGAATGCTCGCTGCCCTGGTCGGCGGTG 1050
 20 G Q S L R S G E C S L A L V G G
 TCAACGGTGATGGCGTCGCCCGGGATTCTCGAGTTCTCCGGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGCGGGGAAGGGCTTCGGCGCGGGCGCGGACGG 1150
 G L A R D G R A K A F G A G A D G
 25 TACGAGCTTCGCCGAGGGCGCCGGTGCCTGGTGGTCGAGCGGGCTCTCG 1200
 T S F A E G A G A L V V E R L S
 ACGCGGAGGCCACGGGACACCCGCTCTCGCCCTCGTACCGGGCTCGCG 1250
 D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCCGTCGAACGGTCTGTCGGCCCGAACGGCCCCTC 1300
 30 A N S D G A S N G L S A P N G P S
 CCAGGAACCGCTCATCACCAGGCCCTCGCGAACCGAAAACCTCACCCCG 1350
 Q E R V I H Q A L A N A K L T P
 CCGATGTCGACCGGGTCGAGGCACGGCACCCGACCCGCTCGCGAC 1400
 A D V D A V E A H G T G T R L G D
 35 CCCATCGAGGCCGAGGGCCTGCTCGCGACGTACGGACAGGACCGGGCGAC 1450
 P I E A Q A L L A T Y G Q D R A T
 GCCCCTGCTGCTCGGCTCGCTGAAGTCGAACATGGGCACGCCAGGCC 1500
 P L L G S L K S N I G H A Q A
 CGTCAGGGTCGCCGGATCATCAAGATGGTCGAGGCCATCCGGCACCGG 1550
 40 A S G V A G I I K M V Q A I R H G
 GAACTGCCGCCACACTGCACGCCGACGAGCCGTCGCCGACGTCGACTG 1600
 E L P P T L H A D E P S P H V D W
 GACGGCCGGTGCCTCGAGCTCTGACGTGGCCCGGGCTGGCCGGGA 1650
 T A G A V E L L T S A R P W P G
 45 CCGGTCGCCCGCGCCGCTGCCGCTCGTCGTTGGCGTGAGCGGCACG 1700
 T G R P R R A A V S S F G V S G T
 AACGCCACATCATCCTTGAGGCAGGACCGGTCAAAACGGGACGGGTCGA 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCGATCGAGGCCAGGACCGGTGAAGTAGGACCGGTGAGGCTG 1800
 50 A G A I E A G P V E V G F V E A
 GAGCGCTCCCCGGCGCCGCGCTCAGCACCGGGCGAACGACCTCCGCTG 1850
 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCTTCCCCGGAGGCACTCGACGAGCACGAGATCGGCGCCT 1900
 L V S A R S P E A L D E Q I G R L
 55 GCGCGCCTATCTCGACACCGGGCCGGCGTCGACCGGGCGACGGCGC 1950
 R A Y L D T G P G V D R A A V A
 AGACACTGGCCCGGCGTACGCACTCACCCACCGGGCGTACTGCTCGGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTCATCGGCCTCCCCCGCGGACCAAGGCCGACGAACCGTCTT 2050
 60 D T V I G A P P A D Q A D E L V F

CGCTCTACTCCGGTCAGGGCACCCAGCATCCCGCATGGGCGAGCAGCTAG 2100
 V Y S G Q G T Q H P A M G E Q L
 CCGCCCGCGTCCCCGTCTCGCGCGATCCATCAGCAGGTGTGGGACCTG 2150
 A A A F P V F A R I H Q Q V W D L
 5 CTCGATGTGCCGATCTGGAGGTGAACGAGACCGGTTACGCCAGCCGGC 2200
 E D V P D L E V N E T G Y A Q P A
 CCTGTTCGCAATGCAGGTGGCTCTGTTGGCTGCTGGAATCGTGGGTG 2250
 L F A M Q V A L F G L L E S W G
 10 TACGACCGGACCGCGGTGATCGGCCATTGGTGGGTGAGCTTGCAGCTGCG 2300
 I R P D A V I G H S V G E L A A A
 TATGTGTCCGGGGTGTGGCTGGAGGATGCCTGCACTTGGTGTGCGC 2350
 Y V S G V W S L E D A C T L V S A
 CGGGGCTCGTCTGATGCAGGCTCTGCCCGGGTGGGTGATGGTGTGCG 2400
 R A R L M Q A L P A G G G V M V A
 15 TCCCGGTCTCGGAGGATGAGGCCCGGGCGTGTGGGTGAGGGTGTGGAG 2450
 I P V S E D E A R A V L G E G V E
 ATCCGGCGGGTCAACGGCCCGTCTGCGGGTGTCTCCGGTGTGAGGC 2500
 I A A V N G P S S V V L S G D E A
 CGCCGTGCTGCAGGCCCGGGCTGGGAAGTGGACGCCGTGGCGA 2550
 20 A V L Q A A E G L G K W T R L A
 CCAGCCACCGCTTCCATTCCGCCGTATGAAACCCATGCTGGAGGAGTTC 2600
 T S H A F H S A R M E P M L E E F
 CGGGCGGTGCGCGAACGGCTGACCTACCGGACGCCGAGGTCTCCATGGC 2650
 R A V A E G L T Y R T P Q V S M A
 25 CGTTGGTGTACAGGTGACCACCGCTGAGTACTGGGTGCGGGCAGGTCCGGG 2700
 V G D Q V T T A E Y W V R Q V R
 ACACGGTCCGGTTCGCGAGCAGGTGGCTCGTACGAGGACGCCGTGTC 2750
 D T V R F G E Q V A S Y E D A V F
 GTCGAGCTGGGTGCCGACCGGTCACTGGCCCGCCTGGTCGACGGTGTGCG 2800
 30 V E L G A D R S L A R L V D G V A
 GATGCTGCACGGCGACCACGAAATCCAGGCCGCGATGGCCCTGGCCC 2850
 M L H G D H E I Q A A I G A L A
 ACCTGTATGTCAACGGCGTACGGTCGACTGGCCCGCCTGGCGAT 2900
 H L Y V N G V T V D W P A L L G D
 35 GCTCCGGCAACACGGGTCTGGACCTTCCGACATACGCCCTCCAGCACCA 2950
 A P A T R V L D L P T Y A F Q H Q
 SCGCTACTGGCTCGAGTCGGCTCCCCCGGCCACGCCACTCGGGCACC 3000
 R Y W L E S A P P A T A D S G H
 CGTCCTCGGACCGGAGTCGCCGTGCCGGTCGCCGGCGGGTGTTC 3050
 40 P V L G T F V A V A G S P G R V F
 ACGGGTCCCGTGCCGCCCGGTGCCGGACCGCGCGGTGTTCATGCCGA 3100
 T G P V P A G A D R A V F I A E L
 GGCCTCGCCGCCGCGACGCCACCGACTGCCACGGTCGAACAGCTCG 3150
 A L A A A D A T D C A T V E Q L
 45 ACGTACCTCCGTGCCCGGGATCCGCCGCCGGCAGGGCACCGCGAG 3200
 D V T S V P G G S A R G R A T A Q
 ACCTGGGTGATGAACCCGCCGCGACGGCGGGCTTCACCGTCCA 3250
 T W V D E P A A D G R R F T V H
 CACCCCGCGTGGCGACGCCCGTGGACGCTGCACGCCGAGGGGTTCTCC 3300
 50 T R V G D A P W T L H A E G V L
 GCCCCGGCCCGTGGCCCGAGCCCGAAGCCGTCGACACCGCCTGGCCCCCG 3350
 R P G R V P Q P E A V D T A W P P
 CGGGCGCGGTGCCGCCGGACGGGCTGCCGGGTGGCGACGCCGGA 3400
 P G A V P A D G L P G A W R R A D
 55 CGAGGTCTCGTCAAGCCGAAGTCGACAGCCCTGACGGCTTCGTGGCAC 3450
 Q V F V E A E V D S P D G F V A
 ACCCCGACCTGCTCGACGCCGGTCTCCGCCGGTGGCGACGCCGCGC 3500
 E F D L L D A V F S A V G D G S R
 CAGCCGACCGGATGGCGCGACCTCGCGGTGCACGCCGTCGGACGCCACCGT 3550
 60 Q P T G W R D L A V H A S D A T V

GCTGCGCGCTGCCTCACCCGCCCGACAGTGGTGTCTGGAGCTCGCCG 3600
 L R A C L T R R D S G V V E L A
 CCTTCGACGGTGCCGGAATGCCGGTGCCTACCGCGGAGTCGGTACGCTG 3650
 A F D G A G M P V L T A E S V T L
 5 GGCAGGGTCCGCGTCCGGCAGGCCGATCCGACGGAGTCGGACGGCTCTGCTCG 3700
 G E V A S A G G S D E S D G L L R
 GCTTGAGTGGTGGCCGGAGGCCACTACGACGGTGCCGACGAGC 3750
 L E W L P V A E A H Y D G A D E
 TGCCTCGAGGGCTACACCCCTCATCACCGCCACACACCCGACGACCCGAC 3800
 10 L P E G Y T L I T A T H F D D P D
 GACCCCAACCAACCCCCACAACACACCCACACGCACCCACACAAACAC 3850
 D P T N P H N T P T R T H T Q T T
 ACCGCTCCTCACCGCCCTCCAACACACCACCTCATCACCAACCAACCC 3900
 R V I T A L Q H H L I T T N H T
 15 TCATCGTCCACACCACCCACCGACCCCCCAGGCGCCGCGTCACCGGCCTC 3950
 L I V H T T T D P P G A A V T G L
 ACCCGCACCCACAAAACGAACACCCCGGCCGATCCACCTCATCGAAC 4000
 T R T A Q N E H P G R I H L I E T
 CCACCAACCCCCACACCCACTCCCCCTCACCCAACCTCACCAACCC 4050
 20 H H P H T P L P L T Q L T T L H
 AACCCCCACCTACGCCCTACCAACAACACCCCTCCACACCCCCCACC 4100
 Q P H L R L T N N T L H T P H L T
 CCCATCACCAACCCACCAACACCCACCAACACCCCCAACACCCCCACC 4150
 P I T T H H N T T T T P N T P P
 25 CCTCAACCCCCAACCGCCATCCTCATCACGGGGCTCCGGCACCCCTCG 4200
 L N P N H A I L I T G G S G T L
 CCGGCATCCTCGCCGCCACCTCAACCACCCCCCACCTACCTCTCTCC 4250
 A G I L A R H L N H P H T Y L L S
 CGCACACCAACCCCCACCAACACCCGGCACCCACATCCCCCTGGGACCT 4300
 30 R T P P P P T T P G T H I P C D L
 CACCGACCCCACCCAAATCACCAAGCCCTCACCCACATACCAACCCCC 4350
 T D P T Q I T Q A L T H I P Q P
 TCACCGGCATCTTCCACACCGCCGCCACCCCTCCGACGACGCCACCCCTCACC 4400
 35 L T G I F H T A A T L D D A T L T
 AACCTCACCCCCCAACACCTCACCAACCCCCCTCAACCCAAAGCCGACGC 4450
 N L T P Q H L T T T L Q P K A D A
 CGCTGGCACCTCCACCAACACCCCCAACCAACCCACTACCCACTTCG 4500
 A W H L H H H T Q N Q P L T H F
 40 TCCTCTACTCCAGCGCCGCCACCCCTCGGCAGCCCCGGCCAAGCCAAC 4550
 V L Y S S A A A T L G S P G Q A N
 TACGCCGCCGCCAACGCCCTCCGACGCCCTGCCACCCACCGGCCACAC 4600
 Y A A A N A F L D A L A T H R H T
 CCAAGGACAACCCGCCACCAACCATGCCCTGGGCATGTGGCACACCAACCA 4650
 Q G Q P A T T I A W G M W H T T
 45 CCACACTCACCAAGCCAACTCACCGACAGCGACCGCGACCGCATCCGCCGC 4700
 T T L T S Q L T D S D R D R I R R
 GGCGCTTCTGCCGATCTGGACGACGAGGGCATGC
 G G F L P I S D D E G M

50 The *NheI-XbaI* hybrid FK-506 PKS module 8 containing the AT domain of module 13 of rapamycin is shown below.

55 GCGATGCCGCTGTACGAGGGCGGCCACGGCGCACCGGAAGTCCCGTGTGGTG 50
 M R L Y E A A R R T G S P V V V
 GCGGCCGCCGCTCGACGACGCCGGACGTGCCGCTGCGCGGGCTGCG 100
 A A A L D D A P D V P L L R G L R
 CGCTACGACCGTCCGGCGTGCCGCCGTCGGGAACGCTCTGCGCCGACCC 150
 R T T V R R A A V R E R S L A D
 GCTCGCCGTGCTGCCGACGACGAGCGCGCCGACGCCCTCCCTCGCGTTCG 200
 R S F C C P T T S A P T F P S R S

TCCCTGGAAACAGCACCGCCACCGTGCTCGGCCACCTGGCGCCGAAGACAT 250
 S W N S T A T V L G H L G A E D I
 CCCGGCGACGACGACGTTCAAGGAACCTCGGCATCGACTCGCTCACCGCGG 300
 F A T T T F K E L G I D S I T A
 5 TCCAGCTGCGCAACCGCGCTGACCCACGGCGACCGCGTACGCCCTAACGCC 350
 V Q L R N A L T T A T G V R L N A
 ACAGCGGTCTCGACTTCCGACGCCGCGCTCGCCCGAGAGACTCGG 400
 T A V F D F P T P R A L A A R L G
 CGACGAGCTGGCCCGTACCCCGCCGCCGCTCGCGCCCGACCGCGCCA 450
 10 D E L A S T R A P V A A R T A A
 CGCGGCCGCCACGACGAACCGCTGGCGATCGTGGCATGGCCTGCCGT 500
 T A A A H D E P L A I V G M A C R
 CTGCGGGCGGGGTCCGCTCGCACAGGAGCTGTGGCGTCTCGCGTC 550
 15 L P G G V A S P Q E L W R L V A S
 CGGCACCGACGCCATCACGGAGTTCCCGCCGACCGCGCTGGGACGTGG 600
 G T D A I T E F P A D R G W D V
 ACCCGCTCTACGACCCGGACCCCGACGCCATCGGCAAGACCTCGTCCGG 650
 D A L Y D P D P D A I G K T F V R
 CACGGCGGCTTCCTCGACGGTGCACCGGCTTCGACCGCGCTTCGG 700
 20 H G G F L D G A T G F D A A F F G
 GATCAGCCCGCGAGGCCCTGGCCATGGACCCCGCAGCAACGGGTGCTCC 750
 I S P R E A L A M D P Q Q R V L
 TGGAGACGTCTGGAGGGCGTTCGAAAGCGCGGGCATCACCCGGACGCG 800
 L E T S W E A F E S A G I T P D A
 25 GCGCGGGGCAGCGACACCGCGTGTTCATCGCGCGTCTCCTACGGGTA 850
 A R G S D T G V F I G A F S Y G Y
 CGGCACGGGTGCGGATACCAACGGCTTCGGCGACAGGGTCGAGACCA 900
 G T G A D T N G F G A T G S Q T
 GCGTGTCTCCGGCCGCTCTCGTACTTCTACGGTCTGGAGGGCCCTCG 950
 30 S V L S G R L S Y F Y G L E G P S
 GTCACGGTCGACACCGCCCTGCTCGTCACTGGTCGCCCTGCACCAAGGC 1000
 V T V D T A C S S S L V A L H Q A
 AGGGCAGTCCTCGCGCTCGGCGAATGCTCGCTCGCCCTGGTCGGCGGTG 1050
 G Q S L R S G E C S L A L V G G
 35 TCAACGGTGTGGCGTCGCCCGGATTCTCGAGTTCTCCCGCAGCGC 1100
 V T V M A S P G G F V E F S R Q R
 GGGCTCGCGCCGGACGGCGGGCGAAGGCCTCGCGCGGGCGGGACGG 1150
 G L A P D G R A K A F G A G A D G
 TACGAGCTTCGCCGAGGGCGCCGGTCCCTGGTGGTCAGCGGGCTCTCG 1200
 40 T S F A E G A G A L V V E R L S
 ACGCGGAGCGCCACGGGCACACCGCTCTCGCCCTCGTACCGGGCTCCGG 1250
 D A E R H G H T V L A L V R G S A
 GCTAACTCCGACGGCGCTCGAACGGTCTGTCGGCGCCGAACGGCCCCCTC 1300
 A N S D G A S N G L S A P N G P S
 45 CCAGGAACCGCTCATCCACCAAGGCCCTCGCGAACCGCAAACCGGAAACTCACCCCG 1350
 Q E R V I H Q A L A N A K L T P
 CCGATGTGCGACGGCGTCAGGGCACGGCACCGGCCCTGGCGAC 1400
 A D V D A V E A H G T G T R L G D
 CCCATCGACCGCGAGGCCCTGCTCGCGACGTACGGACAGGACCGGGCGAC 1450
 50 P I E A C A L L A T Y G Q D R A T
 GCGCCCTGCTCGCTCGCTCGCTGAAGTCGAACATCGGGCACGCCAGGCC 1500
 P L L L G S I K S N I G H A Q A
 CGTCAGGGGTGCGCCGGATCATCAAGATGGTGCAGGCCATCGGCCACGGG 1550
 A S G V A G I I K M V Q A I R H G
 55 GAACTCCCGGCGACACTGCACGGGACGAGCCCTCGCCGACCTCGACTG 1600
 E L P P T L H A D E P S P H V D W
 GAGGGCGGGTGGCGCTCGAGCTCTGACGTGGCCCGGGCGTGGCGGGGA 1650
 T A G A V E L L T S A R P W F G
 CCGTCGCCCGCGCCGCTGCCGTCTCGTCGTTGGCGTGAGCGGCACG 1700
 60 T G R P R R A A V S S F G V S G T

AACGCCACATCATCCTGAGGCAGGACCGGTCAAAACGGGACCGGTGCA 1750
 N A H I I L E A G P V K T G P V E
 GGCAGGAGCGATCGAGGCAGGACCGGTGCAAGTAGGACCGGTGAGGCTG 1800
 A G A I E A G P V E V G P V E A
 5 GACCGCTCCCCGGCGCCGCCGTGCAAGCACCCGGCGAAGACCTCCGCTG 1850
 G P L P A A P P S A P G E D L P L
 CTCGTGTCGGCGCGTCCCCGGAGGCACTCGACGAGCAGATCGGGCGCCT 1900
 L V S A R S P E A L D E Q I G R L
 GCGCGCTATCTGACACCGGCCGGCGTCGACCGGGCGCCGTGGCGC 1950
 10 R A Y L D T G P G V D R A A V A
 AGACACTGGCCGGCGTACGCACTCACCCACCGGGCGTACTGCTCGGG 2000
 Q T L A R R T H F T H R A V L L G
 GACACCGTCATCGGCCTCCCCCGCGGACCAGGCCGACGAACCTCGTCTT 2050
 D T V I G A P P A D Q A D E L V F
 15 CGTCTACTCCGGTCAGGGCACCCAGCATCCCGCGATGGCGAGCAGCTAG 2100
 V Y S G Q G T Q H P A M G E Q L
 CGGATTCTGTCGGTGGTGGCGCCGAGCGGATGGCCGAGTGTGCGGCCG 2150
 A D S S V V F A E R M A E C A A A
 TTGCGCGAGTTCTGACTGGATCTGTCACGGTTCTGGATGATCCGGC 2200
 20 L R E F V D W D L F T V L D D P A
 GGTGGTGGACCGGGTTGATGTGGTCCAGCCCCTCTGGCGATGATGG 2250
 V V D R V D V V Q P A S W A M M
 TTCCCTGGCCGGTGTGGCAGGCAGGCCGGTGTGCGGCCGGATGCGGTG 2300
 V S L A A V W Q A A G V R P D A V
 25 ATCGGCCATTCGCAAGGGTGAGATCGCCGAGCTGTGCGGGTGCCTG 2350
 I G H S Q G E I A A A C V A G A V
 GTCACTACGGATGCCCGGATCGTGCACCTTGCAGCCAGGCGATCG 2400
 S L R D A A R I V T L R S Q A I
 CCCGGGCGCTGGCGGGGGCGATGGCATCCGCGCCCTGCCCCCG 2450
 30 A R G L A G R G A M A S V A L P A
 CAGGATGTGAGCTGGTCGACGGGCTGGATCGCCGCCACAACGGGCC 2500
 Q D V E L V D G A W I A A H N G P
 CGCCTCCACCGTGATCGCGGGCACCCCGGAAGCGGTGACCATGTCCTCA 2550
 A S T V I A G T P E A V D H V L
 35 CCGCTCATGAGGCACAAGGGTGCAGGGGATCGCGGACTACCGTCGACTAT 2600
 T A H E A Q G V R V R R I T V D Y
 GCCTCGCACACCCCGCACCTCGAGCTGATCCGCGACGAACACTCGACAT 2650
 A S H T P H V E L I R D E L L D I
 CACTAGCGACAGCAGCTCGCAGACCCCGCTCGTGCCTGGTGTGACCG 2700
 40 T S D S S S Q T P L V P W L S T
 TGGACGGCACCTGGTGCACAGCCCCTGGACGGGGAGTACTGGTACCGG 2750
 V D G T W V D S P L D G E Y W Y R
 AACCTGCGTGAACCGGTGGTTCCACCCCGCGTCAGCCAGTTGCAGGC 2800
 N L R E P V G F H P A V S Q L Q A
 45 CCAGGGCGACACCGTGGTCGAGGTCAAGGCCAGCCCGGTGTGTTGC 2850
 Q G D T V F V E V S A S P V L L
 AGGCGATGGACGACGATGTCGTACGGTTGCCACGCTGCGTGTGACGAC 2900
 Q A M D D D V V T V A T L R R D D
 GGCACGCCACCCGGATGCTCACCGCCCTGGCACAGGCCTATGTCACGG 2950
 50 G D A T R M L T A L A Q A Y V H G
 CCTCACCGTCGACTGGCCGCCATCCTCGGACCAACACCCGGTAC 3000
 V T V D W P A I L G T T T R V
 TGGACCTTCCGACCTACGCGCTTCAACCCAGCGTACTGGCTCGAGTCG 3050
 L D L P T Y A F Q H Q R Y W L E S
 55 GCTCCCCCGGCCACGGCGACTCGGGCCACCCCGTCTCGGCACCGGAGT 3100
 A P P A T A D S G H P V L G T G V
 CGCGCTGGCGGGTGCACGGGGCCGGTGTACCGGTCCCCGTGCCCGCCG 3150
 A V A G S P G R V F T G P V P A
 GTGCGGACCCGCGCGGTGTTCATGCCGAACGGCGCTGCCGCCGAC 3200
 60 G A D R A V F I A E L A L A A A D

5 GCCACCGACTGCGC CACGGTCGAACAGCTCGACGTACCTCCGTGCCCGG 3250
A T D C A T V E Q L D V T S V P G
CGGATCCGCCCGCAGGGC CACCGCGCAGACCTGGGTCGATGAACCCG 3300
S S A R G R A T A Q T W V D E P
CGGCCGACGGCGCCCTCACCGTCCACACCCCGCGTCGGCGACGCC 3350
A A D G R R R F T V H T R V G D A
CGTGGACGCTGCACGCCGAGGGGTTCTCCGCCCGGCCGCGTGCCCCA 3400
P W T L H A E G V L R P G R V P Q
GCCCGAAGCCGTCGACACCGCCTGGCCCCCGCCGGCGCGGTGCCGCCG 3450
10 P E A V D T A W P P P G A V P A
ACGGGCTGCCCGGGCGTGGCGACCGCGGGACCAGGTCTTCGTCGAAGCC 3500
D G L P G A W R R A D Q V F V E A
GAGTCGACAGCCCTGACGGCTTCGTGGCACACCCGACCTGCTCGACGC 3550
E V D S P D G F V A H P D L L D A
33TCTTCTCCCGCGTGGCGACGGGAGGCCAGCCGACCGGATGGCGCG 3600
V F S A V G D G S R Q P T G W R
ACCTCGCGGTGCACCGCTCGGACGCCACCGTGCTCGCGCCTGCCCTACCC 3650
D L A V H A S D A T V L R A C L T
GCCCGCGACAGTGGTGTGCTGGAGCTCGCCGCCCTCGACGGTGGCAAT 3700
20 R R D S G V V E L A A F D G A G M
GCCGGTGTCTACCGCGGAGTCGGTGACGCTGGCGAGGTGCGTCGGCAG 3750
P V L T A E S V T L G E V A S A
GCGGATCCGACGAGTCGGACGGTCTGCTGGCTTGAGTGGTTGCCGGTG 3800
3 G S D E S D G L L R L E W L P V
GCGGAGGCCCACTACGACGGTGCCGACGAGCTGCCGAGGGCTACACCCCT 3850
A E A H Y D G A D E L P E G Y T L
CATCACCGCCACACACCCGACGACCCGACGACCCACCAACCCCCACA 3900
I T A T H P D D P D D P T N P H
ACACACCCACACGCAACACACACAAACCACACGCGTCCTCACCGCCCTC 3950
30 N T P T R T H T Q T T R V L T A L
CAACACCCACCTCATACCAACCAACCACCCCTCATCGTCCACACCCAC 4000
Q H H L I T T N H T L I V H T T T
CGACCCCCCAGGGCGCCCGTACCGGCTCACCCGACCGCACAAACCG 4050
D P P G A A V T G L T R T A Q N
35 AACACCCCCGGCGCATCCACCTCATCGAAACCCACACCCCCACACCCCA 4100
E H P G R I H L I E T H H P T P
CTCCCCCTCACCAACTCACCAACCCCTCCACCAACCCACCTACGGCTCAC 4150
I P L T Q L T T L H Q P H L R L T
CAACAACACCCCTCACACCCCCCACCTCACCCCCATCACCAACCCACCA 4200
40 N N T L H T P H L T P I T T H H
ACACCAACACAACACCCCCAACACCCCCACCCCTCAACCCCAACCCACGCC 4250
N T T T T P N T P P L N P N H A
ATCCTCATCACGGGGCTCCGGCACCCCTCGCCGGATCCTCGCCGCCA 4300
I L I T G G S G T L A G I L A R H
45 CCTCAACCAACCCCCCACCTACCTCCCTCCCGCACACCAACCCCCCA 4350
L N H P H T Y L L S R T P P P P
CCACACCCGGCACCCACATCCCCCTCGACCTCACCGACCCACCCAAATC 4400
T T P G T H I P C D L T D P T Q I
ACCCAAGCCCTCACCCACATACCAACACCCCTCACCGGATCTCCACAC 4450
50 T Q A L T H I P Q P L T G I F H T
CGCCGCCACCCCTCGACGCCACCCCTCACCAACCTCACCCCCAACACC 4500
A A T L D D A T L T N L T P Q H
TCACCAACCAACCCCTCACCAACCCAAAGCCGACGCCGCTGGCACCTCCACCA 4550
L T T T L Q P K A D A A A W H L H H
CACACCCAAAACCAACCCCTCACCCACTTCGTCCTCTACTCCAGCGCCGC 4600
H T Q N Q P L T H F V L Y S S A A
CCCGACCCCTCGGCACCCCCGGCCAAGCCAACTACGCCGCCAACCGCCT 4650
A T L G S P G Q A N Y A A A N A
TCCTCGACGCCCTCGCCACCCACCGGCCACACCCAAAGGACAACCCGCCACC 4700
60 F L D A L A T H R H T Q G Q P A T

ACCATCGCTGGGCATCTGGCACACCACCACTCACCAACT 4750
 T I A W G M W H T T T T L T S Q L
 CACCGACAGCGACCGCGACCGCATCCGCCGGCGGCTTCCTGCCGATCT 4800
 T D S I R D R I R R G G F I P I
 5 CGGACGACAGGGCATGC
 S D D E G M

Example 3

Recombinant PKS Genes for 13-desmethoxy FK-506 and FK-520

10 The present invention provides a variety of recombinant PKS genes in addition to those described in Examples 1 and 2 for producing 13-desmethoxy FK-506 and FK-520 compounds. This Example provides the construction protocols for recombinant FK-520 and FK-506 (from *Streptomyces* sp. MA6858 (ATCC 55098), described in U.S. Patent Nos. 5,116,756, incorporated herein by reference) PKS genes in which the module 8 AT coding
 15 sequences have been replaced by either the *rapAT3* (the AT domain from module 3 of the rapamycin PKS), *rapAT12*, *eryAT1* (the AT domain from module 1 of the erythromycin (DEBS) PKS), or *eryAT2* coding sequences. Each of these constructs provides a PKS that produces the 13-desmethoxy-13-methyl derivative, except for the *rapAT12* replacement, which provides the 13-desmethoxy derivative, i.e., it has a hydrogen where the other
 20 derivatives have methyl.

Figure 7 shows the process used to generate the AT replacement constructs. First, a fragment of ~4.5 kb containing module 8 coding sequences from the FK-520 cluster of ATCC 14891 was cloned using the convenient restriction sites *SacI* and *SphI* (Step A in Figure 7). The choice of restriction sites used to clone a 4.0 - 4.5 kb fragment comprising module 8 coding sequences from other FK-520 or FK-506 clusters can be different depending on the DNA sequence, but the overall scheme is identical. The unique *SacI* and *SphI* restriction sites at the ends of the FK-520 module 8 fragment were then changed to unique *Bgl* II and *Nsi* I sites by ligation to synthetic linkers (described in the preceding Examples, see Step B of Figure 7). Fragments containing sequences 5' and 3' of the AT8 sequences were then amplified using primers, described above, that introduced either an *Avr* II site or an *Nhe* I site at two different KS/AT boundaries and an *Xba* I site at the AT/DH boundary (Step C of Figure 7). Heterologous AT domains from the rapamycin and erythromycin gene clusters were amplified using primers, as described above, that introduced the same sites as just described (Step D of Figure 7). The fragments were ligated to give hybrid modules with in-frame fusions at the KS/AT and AT/DH boundaries (Step E of Figure 7). Finally, these hybrid modules were ligated into the *Bam* HI and *Pst* I sites of the

KC515 vector. The resulting recombinant phage were used to transform the FK-506 and FK-520 producer strains to yield the desired recombinant cells, as described in the preceding Examples.

The following table shows the location and sequences surrounding the engineered site of each of the heterologous AT domains employed. The FK-506 hybrid construct was used as a control for the FK-520 recombinant cells produced, and a similar FK-520 hybrid construct was used as a control for the FK-506 recombinant cells.

Heterologous AT	Enzyme	Location of Engineered Site
FK-506 AT8 (hydroxymalonyl)	<i>AvrII</i>	GGCCGT <u>ccgcgc</u> CGTGC GGCGGT CTCGTCGTT G R P R R A A V S S F
	<i>NheI</i>	ACCCAGCAT <u>CCCGCG</u> ATGGGTGAGCG <u>gatcgac</u> T Q H P A M G E R L A
	<i>XbaI</i>	TACGCCTTCCAGCGGCC <u>TACTGG</u> <u>atcgag</u> Y A F Q R R P Y W I E
rapamycin AT3 (methylmalonyl)	<i>AvrII</i>	GACCGG <u>ccccat</u> CGGGCGGGCGTGTGTCCTTC D R P R R A G V S S F
	<i>NheI</i>	TGGCAGTGGCTGGGGATGGCAGTGC <u>ccatcgac</u> W Q W L G M G S A L R
	<i>XbaI</i>	TACGCCTTCCA <u>ACACCA</u> CAGCGGTACTGG <u>atcgag</u> Y A F Q H Q R Y W V E
rapamycin AT12 (malonyl)	<i>AvrII</i>	GGCCGA <u>acgcgc</u> CGGGCAGGC <u>GTGTGTC</u> CTTC G R A R R A G V S S F
	<i>NheI</i>	TCGCAGCGTGTGG <u>CATGGGTGAGGA</u> <u>actggc</u> S Q R A G M G E E L A
	<i>XbaI</i>	TACGCCTTCCAGCAC <u>CCAGCGC</u> TACTGG <u>atcgag</u> Y A F Q H Q R Y W L E
DEBS AT1 (methylmalonyl)	<i>AvrII</i>	GCGCG <u>Accacac</u> CGGGCGGGGT <u>CTCGTC</u> GTTC A R P R R A G V S S F
	<i>NheI</i>	TGGCAGTGG <u>GGGG</u> CATGG <u>CCGTG</u> <u>Accatcgac</u> W Q W A G M A V D L L
	<i>XbaI</i>	TACCCGTTCCAGCGCG <u>AGCGC</u> GTCTGG <u>atcgaa</u> Y P F Q R E R V W L E
DEBS AT2 (methylmalonyl)	<i>AvrII</i>	GACGGG <u>atcgac</u> CGGGCAGGTGT <u>GTGCG</u> GTTC D G V R R A G V S A F
	<i>NheI</i>	GCCCAGTGG <u>GAAGG</u> CATGG <u>CGCGGG</u> <u>Agttgt</u> A Q W E G M A R E L L
	<i>XbaI</i>	TATCCTTCCAGGG <u>CAAGC</u> GGTCTGG <u>atactg</u> Y P F Q G K R F W L L

The sequences shown below provide the location of the KS/AT boundaries chosen in the FK-520 module 8 coding sequences. Regions where *AvrII* and *NheI* sites were engineered are indicated by lower case and underlining.

5 CCGGCGCCGTCGAACGTGCTGACGTGGCCCGGCCGTGGCCCGAGACCGACCGGGccacagC
 A G A V E L L T S A R P W P E T D R P R
 GTGCCGCCGTCCTCGTTGGGGTGAGCGGCACCAACGCCACGTATCCTGGAGGCCG
 P A A V S S F G V S G T N A H V I L E A
 GACCGCTAACGGAGACGCCCGCGGCATGCCCTCCGGTGACCTTCCCTGCTGGTGTGG
 G P V T E T P A A S P S G D L P L L V S
 CACGCTCACCGGAAGCGCTCGACGAGCAGATCCGCCACTGCCGCCTACCTGGACACCA
 A R S P E A L D E Q I R R L R A Y L D T
 10 CCCCCGACGTCGACCGGGTGGCCGTGGCACAGACGCTGGCCGGCACACACTTCGCC
 T P D V D R V A V A Q T L A R R T H F A
 ACCGCGCCGTCGCTCGGTGACACCGTCATCACCAACACCCCCCGCGACCGGCCGACG
 H R A V L L G D T V I T T P P A D R P D
 AACTCGTCTTCGTCTACTCCGGCCAGGGCACCCAGCATCCCGCGATGGCGAGCAactcg
 E L V F V Y S G Q G T Q H P A M G E Q L
 15 CGCCGCCCATCCCGTGTTCGCCGACGCCCTGGCATGAAGCGCTCCGCCCTTGACAACC
 A A A H P V F A D A W H E A L R R L D N

The sequences shown below provide the location of the AT/DH boundary chosen in
 the FK-520 module 8 coding sequences. The region where an *Xba*I site was engineered is
 20 indicated by lower case and underlining.
 25

TCCTCGGGCTGGGTACGGCACGACGCGATGTGCCCGGTACGCGTTCAAACGGCGGC
 L G A G S R H D A D V P A Y A F Q R R
 ACTACTGGatcgTCGGCACGCCGGCGCATCCGACGCCGGCACCCGTGCTGGGCT
 H Y W I E S A R P A A S D A G H P V L G

The sequences shown below provide the location of the KS/AT boundaries chosen
 in the FK-506 module 8 coding sequences. Regions where *Avr*II and *Nhe*I sites were
 25 engineered are indicated by lower case and underlining.
 30

TCGGCCAGGCCGTGGCCGGACCGCCGTccgcgcCGTGCGGCGTCTCGTCGTTGGG
 S A R P W P R T G R P R R A A V S S F G
 GTGAGCGCACCAACGCCACATCATCCTGGAGGCCGACCCGACCAGGAGGAGCCGTG
 V S G T N A H I L E A G P D Q E E P S
 GCAGAACCGCCGGTGACCTCCCGCTGCTCGTGGCACGGTCCCGGAGGCACGGAC
 A E P A G D L P L L V S A R S P E A L D
 35 GAGCAGATCGGGCGCCCTGCGCGACTATCTGACGCCGCCCCCGCGTGGACCTGGCGGCC
 E Q I G R L R D Y L D A A P G V D L A A
 GTGGCGCGGACACTGGCACCGCGTACGCACTTCTCCACCGCGCGTACTGCTCGGTGAC
 V A R T L A T R T H F S H R A V L L G D
 ACCGTATCACCGCTCCCCCGTGGAACAGCCGGGAGCTCGTCTCGTACTCGGGA
 40 T V I T A P P V E Q P G E L V F V Y S G
 CAGGGCACCCAGCATCCCGCGATGGGTGAGCGactcgCGCAGCCTTCCCGTGTGGCC
 Q G T Q H P A M G E R L A A A A F P V F A
 GACCCGGACGTACCCGCCCTACGCCCTCCAGCGGCCACTGGATCGAGTCGCCGCC
 D P D V P A Y A F Q R R P Y W I E S A P
 45

The sequences shown below provide the location of the AT/DH boundary chosen in
 the FK-506 module 8 coding sequences. The region where an *Xba*I site was engineered is
 indicated by lower case and underlining.
 50

GACCCGGACGTACCCGCCCTACGCCCTCCAGCGGCCACTGGatcgTCCGCC
 P D V P A Y A F Q R R P Y W I E S A P

Example 4

SUBSTITUTE SHEET (RULE 26)

Replacement of Methoxyl with Hydrogen or Methyl at C-15 of FK-506 and FK-520

The methods and reagents of the present invention also provide novel FK-506 and FK-520 derivatives in which the methoxy group at C-15 is replaced by a hydrogen or methyl. These derivatives are produced in recombinant host cells of the invention that express recombinant PKS enzymes the produce the derivatives. These recombinant PKS enzymes are prepared in accordance with the methodology of Examples 1 and 2, with the exception that AT domain of module 7, instead of module 8, is replaced. Moreover, the present invention provides recombinant PKS enzymes in which the AT domains of both modules 7 and 8 have been changed. The table below summarizes the various compounds provided by the present invention.

	Compound	C-13	C-15	Derivative Provided
	FK-506	hydrogen	hydrogen	13, 15-didesmethoxy-FK-506
	FK-506	hydrogen	methoxy	13-desmethoxy-FK-506
15	FK-506	hydrogen	methyl	13,15-didesmethoxy-15-methyl-FK-506
	FK-506	methoxy	hydrogen	15-desmethoxy-FK-506
	FK-506	methoxy	methoxy	Original Compound -- FK-506
	FK-506	methoxy	methyl	15-desmethoxy-15-methyl-FK-506
	FK-506	methyl	hydrogen	13,15-didesmethoxy-13-methyl-FK-506
20	FK-506	methyl	methoxy	13-desmethoxy-13-methyl-FK-506
	FK-506	methyl	methyl	13,15-didesmethoxy-13,15-dimethyl-FK-506
	FK-520	hydrogen	hydrogen	13, 15-didesmethoxy FK-520
	FK-520	hydrogen	methoxy	13-desmethoxy FK-520
	FK-520	hydrogen	methyl	13,15-didesmethoxy-15-methyl-FK-520
25	FK-520	methoxy	hydrogen	15-desmethoxy-FK-520
	FK-520	methoxy	methoxy	Original Compound -- FK-520
	FK-520	methoxy	methyl	15-desmethoxy-15-methyl-FK-520
	FK-520	methyl	hydrogen	13,15-didesmethoxy-13-methyl-FK-520
	FK-520	methyl	methoxy	13-desmethoxy-13-methyl-FK-520
30	FK-520	methyl	methyl	13,15-didesmethoxy-13,15-dimethyl-FK-520

Example 5Replacement of Methoxyl with Ethyl at C-13 and/or C-15 of FK-506 and FK-520

The present invention also provides novel FK-506 and FK-520 derivative compounds in which the methoxy groups at either or both the C-13 and C-15 positions are instead ethyl groups. These compounds are produced by novel PKS enzymes of the invention in which the AT domains of modules 8 and/or 7 are converted to ethylmalonyl specific AT domains by modification of the PKS gene that encodes the module. Ethylmalonyl specific AT domain coding sequences can be obtained from, for example, the FK-520 PKS genes, the niddamycin PKS genes, and the tylosin PKS genes. The novel PKS genes of the invention include not only those in which either or both of the AT domains of modules 7 and 8 have been converted to ethylmalonyl specific AT domains but also those in which one of the modules is converted to an ethylmalonyl specific AT domain and the other is converted to a malonyl specific or a methylmalonyl specific AT domain.

Example 6

Neurotrophic Compounds

15 The compounds described in Examples 1 - 4, inclusive have immunosuppressant activity and can be employed as immunosuppressants in a manner and in formulations similar to those employed for FK-506. The compounds of the invention are generally effective for the prevention of organ rejection in patients receiving organ transplants and in particular can be used for immunosuppression following orthotopic liver transplantation.

20 These compounds also have pharmacokinetic properties and metabolism that are more advantageous for certain applications relative to those of FK-506 or FK-520. These compounds are also neurotrophic; however, for use as neurotrophins, it is desirable to modify the compounds to diminish or abolish their immunosuppressant activity. This can be readily accomplished by hydroxylating the compounds at the C-18 position using

25 established chemical methodology or novel FK-520 PKS genes provided by the present invention.

Thus, in one aspect, the present invention provides a method for stimulating nerve growth that comprises administering a therapeutically effective dose of 18-hydroxy-FK-520. In another embodiment, the compound administered is a C-18,20-dihydroxy-FK-520 derivative. In another embodiment, the compound administered is a C-13-desmethoxy and/or C-15-desmethoxy 18-hydroxy-FK-520 derivative. In another embodiment, the compound administered is a C-13-desmethoxy and/or C-15-desmethoxy 18,20-dihydroxy-FK-520 derivative. In other embodiments, the compounds are the corresponding analogs of

FK-506. The 18-hydroxy compounds of the invention can be prepared chemically, as described in U.S. Patent No. 5,189,042, incorporated herein by reference, or by fermentation of a recombinant host cell provided by the present invention that expresses a recombinant PKS in which the module 5 DH domain has been deleted or rendered non-functional.

The chemical methodology is as follows. A compound of the invention (~200 mg) is dissolved in 3 mL of dry methylene chloride and added to 45 μ L of 2,6-lutidine, and the mixture stirred at room temperature. After 10 minutes, tert-butyldimethylsilyl trifluoromethanesulfonate (64 μ L) is added by syringe. After 15 minutes, the reaction 10 mixture is diluted with ethyl acetate, washed with saturated bicarbonate, washed with brine, and the organic phase dried over magnesium sulfate. Removal of solvent *in vacuo* and flash chromatography on silica gel (ethyl acetate: hexane (1:2) plus 1% methanol) gives the protected compound, which is dissolved in 95% ethanol (2.2 mL) and to which is added 53 μ L of pyridine, followed by selenium dioxide (58 mg). The flask is fitted with a water 15 condenser and heated to 70°C on a mantle. After 20 hours, the mixture is cooled to room temperature, filtered through diatomaceous earth, and the filtrate poured into a saturated sodium bicarbonate solution. This is extracted with ethyl acetate, and the organic phase is washed with brine and dried over magnesium sulfate. The solution is concentrated and purified by flash chromatography on silica gel (ethyl acetate: hexane (1:2) plus 1% 20 methanol) to give the protected 18-hydroxy compound. This compound is dissolved in acetonitrile and treated with aqueous HF to remove the protecting groups. After dilution with ethyl acetate, the mixture is washed with saturated bicarbonate and brine, dried over magnesium sulfate, filtered, and evaporated to yield the 18-hydroxy compound. Thus, the present invention provides the C-18-hydroxyl derivatives of the compounds described in 25 Examples 1 - 4.

Those of skill in the art will recognize that other suitable chemical procedures can be used to prepare the novel 18-hydroxy compounds of the invention. See, e.g., Kawai *et al.*, Jan. 1993, Structure-activity profiles of macrolactam immunosuppressant FK-506 analogues, *FEBS Letters* 316(2): 107-113, incorporated herein by reference. These methods 30 can be used to prepare both the C18-[*S*]-OH and C18-[*R*]-OH enantiomers, with the *R* enantiomer showing a somewhat lower IC₅₀, which may be preferred in some applications. See Kawai *et al.*, *supra*. Another preferred protocol is described in Umbreit and Sharpless, 1977, *JACS* 99(16): 1526-28, although it may be preferable to use 30 equivalents each of

SeO₂ and t-BuOOH rather than the 0.02 and 3-4 equivalents, respectively, described in that reference.

All scientific and patent publications referenced herein are hereby incorporated by reference. The invention having now been described by way of written description and example, those of skill in the art will recognize that the invention can be practiced in a variety of embodiments, that the foregoing description and example is for purposes of illustration and not limitation of the following claims.

Claims

1. An isolated nucleic acid that encodes a CoA ligase, a non-ribosomal peptide synthetase, or a domain of an extender module of a polyketide synthase enzyme that synthesizes FK-520.
5
2. The isolated nucleic acid of claim 1 that encodes an extender module, said module comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.
10
3. The isolated nucleic acid of claim 1 that encodes an open reading frame, said open reading frame comprising coding sequences for two or more extender modules, each extender module comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.
15
4. The isolated nucleic acid of claim 1 that encodes a gene cluster, said gene cluster comprising two or more open reading frames, each of said open reading frames comprising coding sequences for two or more extender modules, each of said extender modules comprising a ketosynthase domain, an acyl transferase domain, and an acyl carrier protein domain.
20
5. The isolated nucleic acid of claim 2, wherein at least one of said domains is a domain of a module of a non-FK-520 polyketide synthase.
25
6. The isolated nucleic acid of claim 1, wherein said nucleic acid is a recombinant vector capable of replication in or integration into the chromosome of a host cell.
30
7. The isolated nucleic acid of claim 6 that is selected from the group consisting of cosmid pKOS034-120, cosmid pKOS034-124, cosmid pKOS065-M27, and cosmid pKOS065-M21.
8. The isolated nucleic acid of claim 5, wherein said non-FK-520 polyketide synthase is rapamycin polyketide synthase, FK-506 polyketide synthase, or erythromycin polyketide synthase.

9. A method of preparing a polyketide, said method comprising transforming a host cell with a recombinant DNA vector of claim 6, and culturing said host cell under conditions such that said polyketide synthase is produced and catalyzes synthesis of said polyketide.

5 10. The method of claim 9, wherein said host cell is a *Streptomyces* host cell.

11. The method of claim 9, wherein said polyketide is selected from the group 10 consisting of FK-520, 13-desmethoxy-FK-520, and 13-desmethoxy-FK-506.

12. A recombinant host cell that expresses a recombinant polyketide synthase selected from the group consisting of: (i) an FK-520 polyketide synthase in which at least one AT domain is replaced by an AT domain of a non-FK-520 polyketide synthase; (ii) an 15 FK-506 polyketide synthase in which at least one AT domain is replaced by an AT domain of a non-FK-506 polyketide synthase; (iii) an FK-520 polyketide synthase in which at least one DH domain has been deleted; (iv) an FK-506 polyketide synthase in which at least one DH domain has been deleted.

20 13. The recombinant host cell of claim 12 that expresses an FK-520 polyketide synthase in which an AT domain of module 8 has been replaced by an AT domain that binds malonyl CoA, methylmalonyl CoA, or ethylmalonyl CoA.

25 14. The recombinant host cell of claim 12 that expresses an FK-506 polyketide synthase in which an AT domain of module 8 has been replaced by an AT domain that binds malonyl CoA, methylmalonyl CoA, or ethylmalonyl CoA.

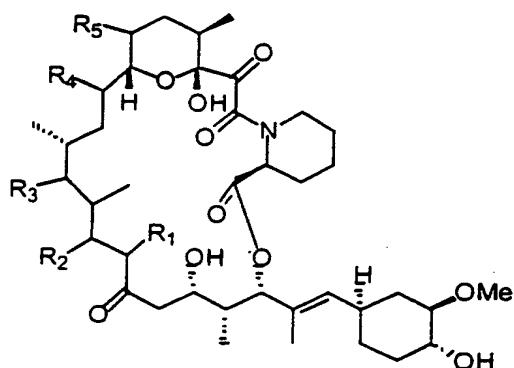
30 15. The recombinant host cell of claim 13, wherein a DH domain of module 5 or module 6 has been deleted.

16. The recombinant host cell of claim 14, wherein a DH domain of module 5 or module 6 has been deleted.

17. A recombinant host cell that comprises recombinant genes coding for enzymes sufficient for synthesis of ethylmalonyl CoA or 2-hydroxymalonyl CoA.

18. A polyketide having the structure

5



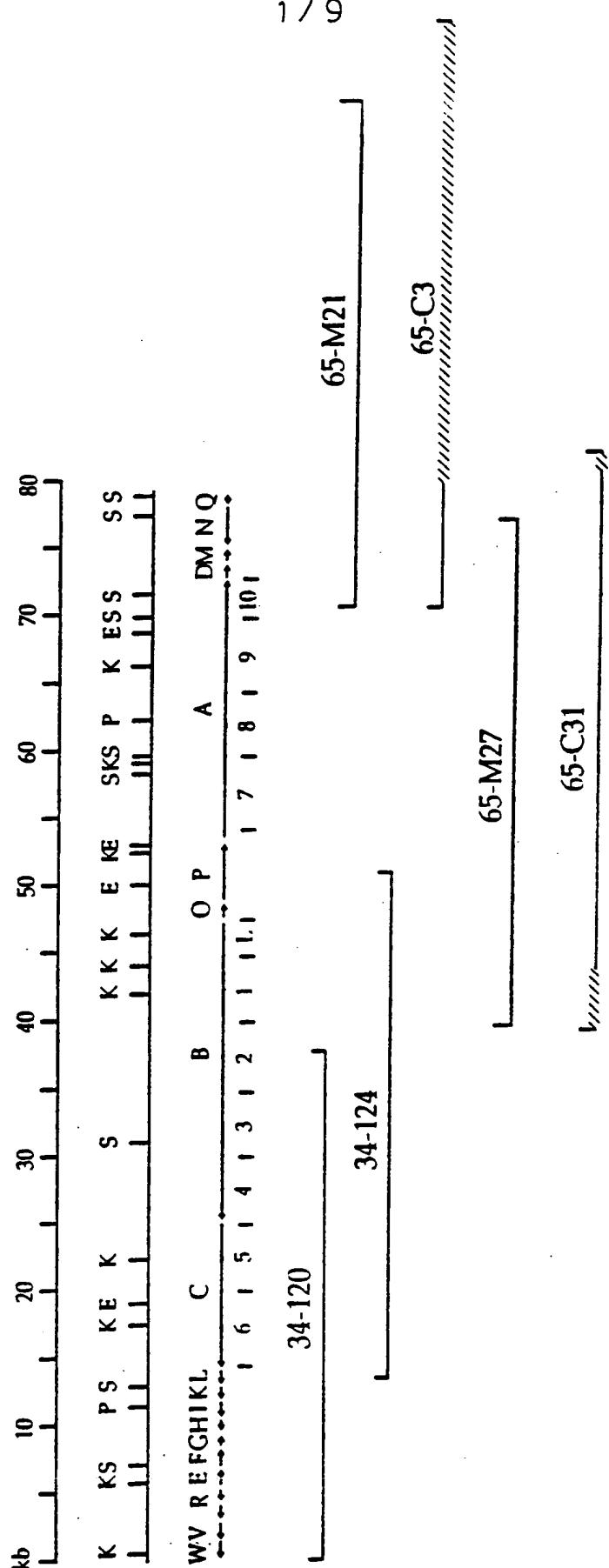
wherein, R₁ is hydrogen, methyl, ethyl, or allyl; R₂ is hydrogen or hydroxyl, provided that when R₂ is hydrogen, there is a double bond between C-20 and C-19; R₃ is hydrogen or 10 hydroxyl; R₄ is methoxyl, hydrogen, methyl, or ethyl; and R₅ is methoxyl, hydrogen, methyl, or ethyl; but not including FK-506, FK-520, 18-hydroxy-FK-520, and 18-hydroxy-FK-506.

15

19. The polyketide of claim 18 that is 13-desmethoxy-FK-506.

20. The polyketide of claim 18 that is 13-desmethoxy-18-hydroxy-FK-520.

1 / 9



SUBSTITUTE SHEET (RULE 26)

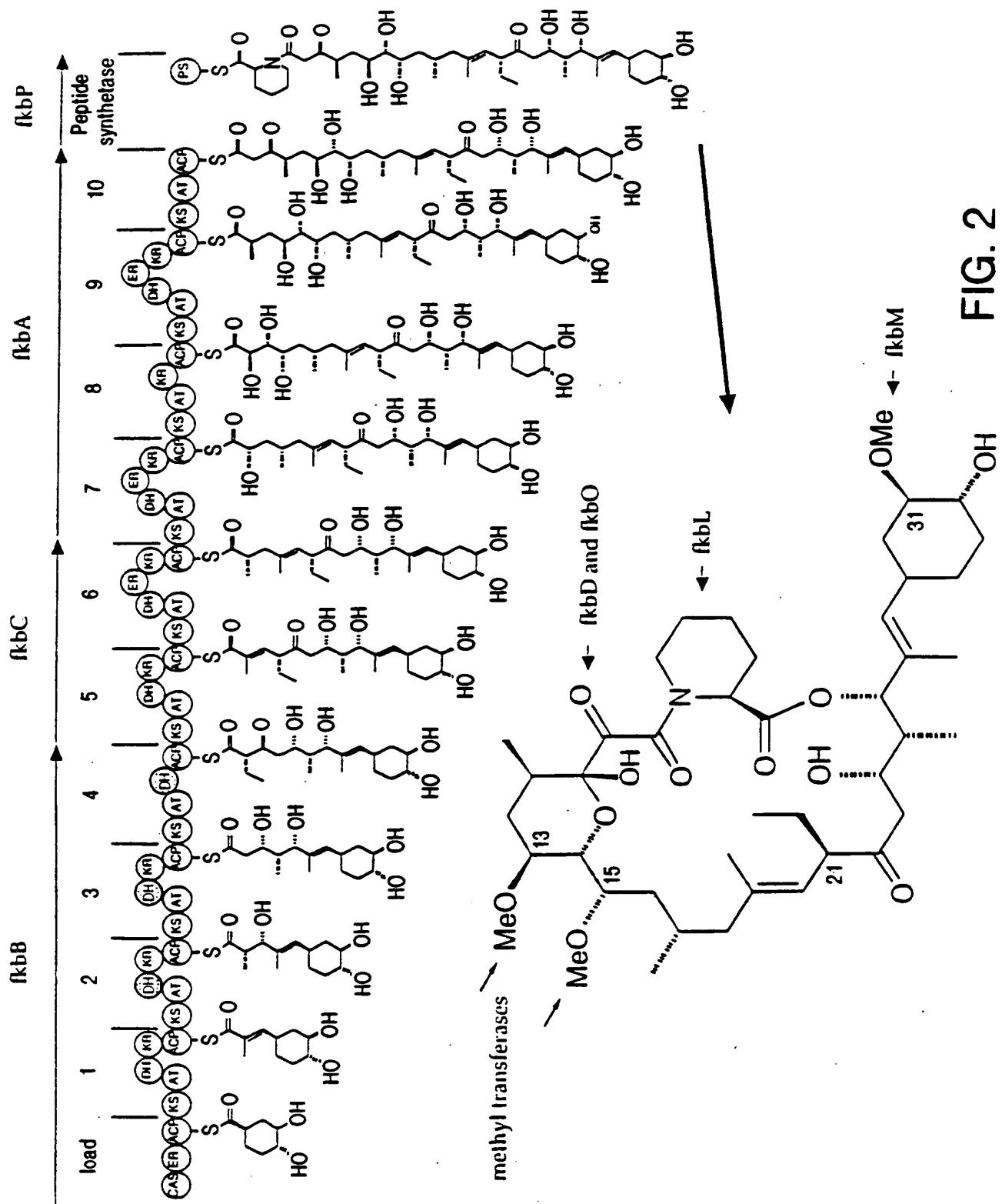


FIG. 2

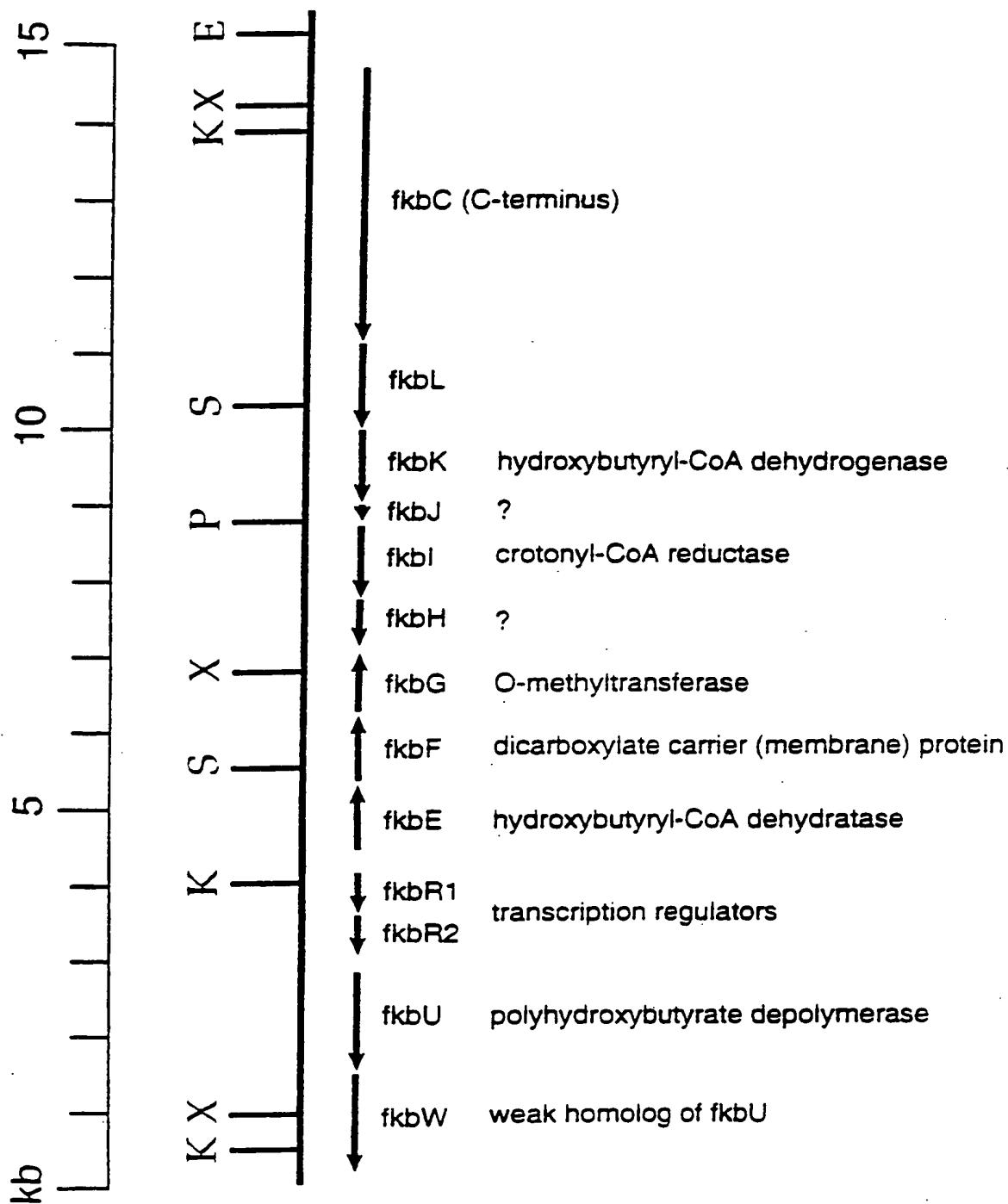


FIG. 3

4 / 9

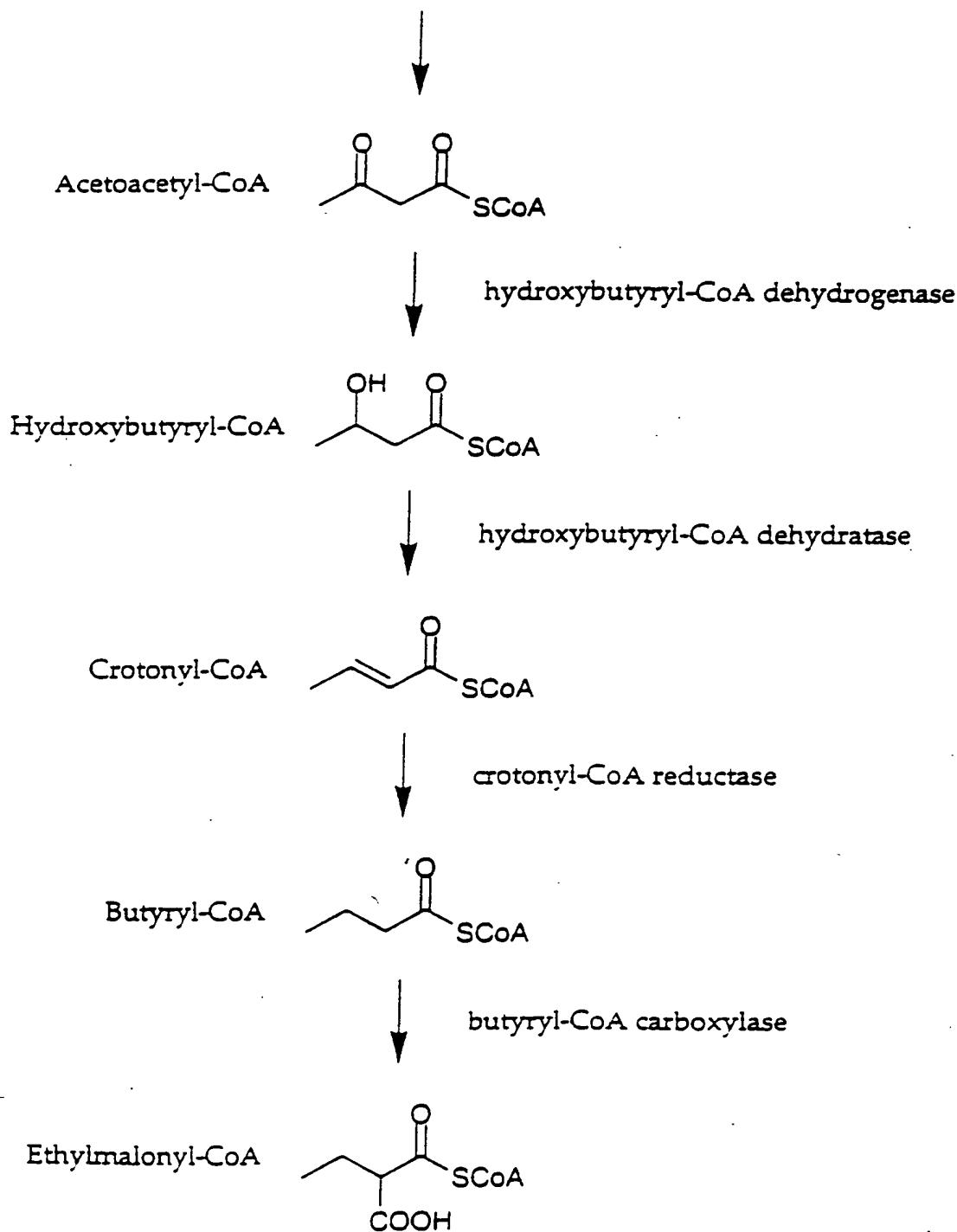


FIG. 4

SUBSTITUTE SHEET (RULE 26)

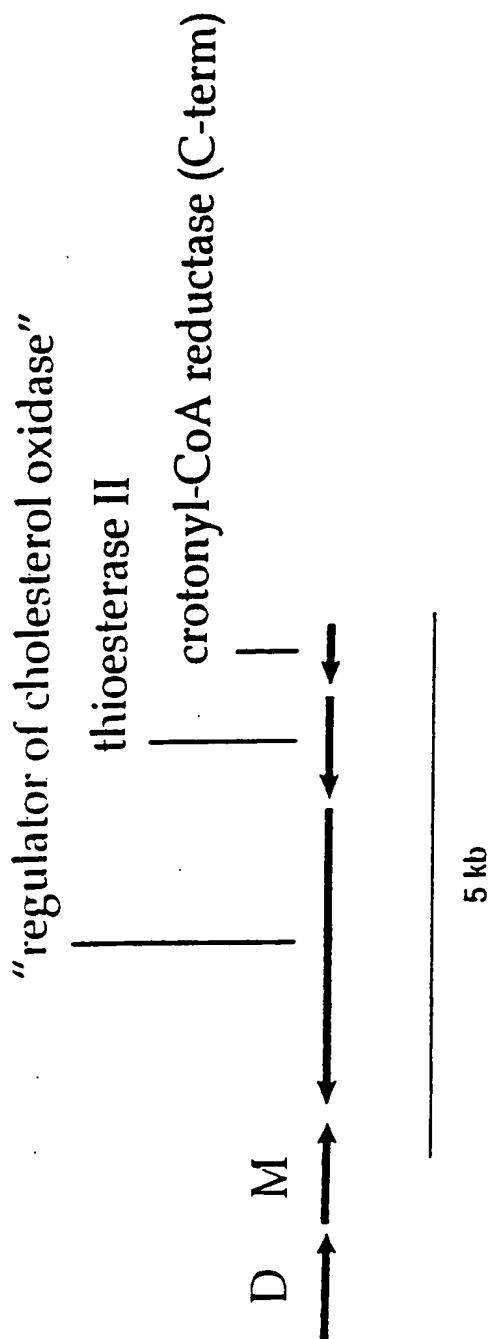


FIG. 5

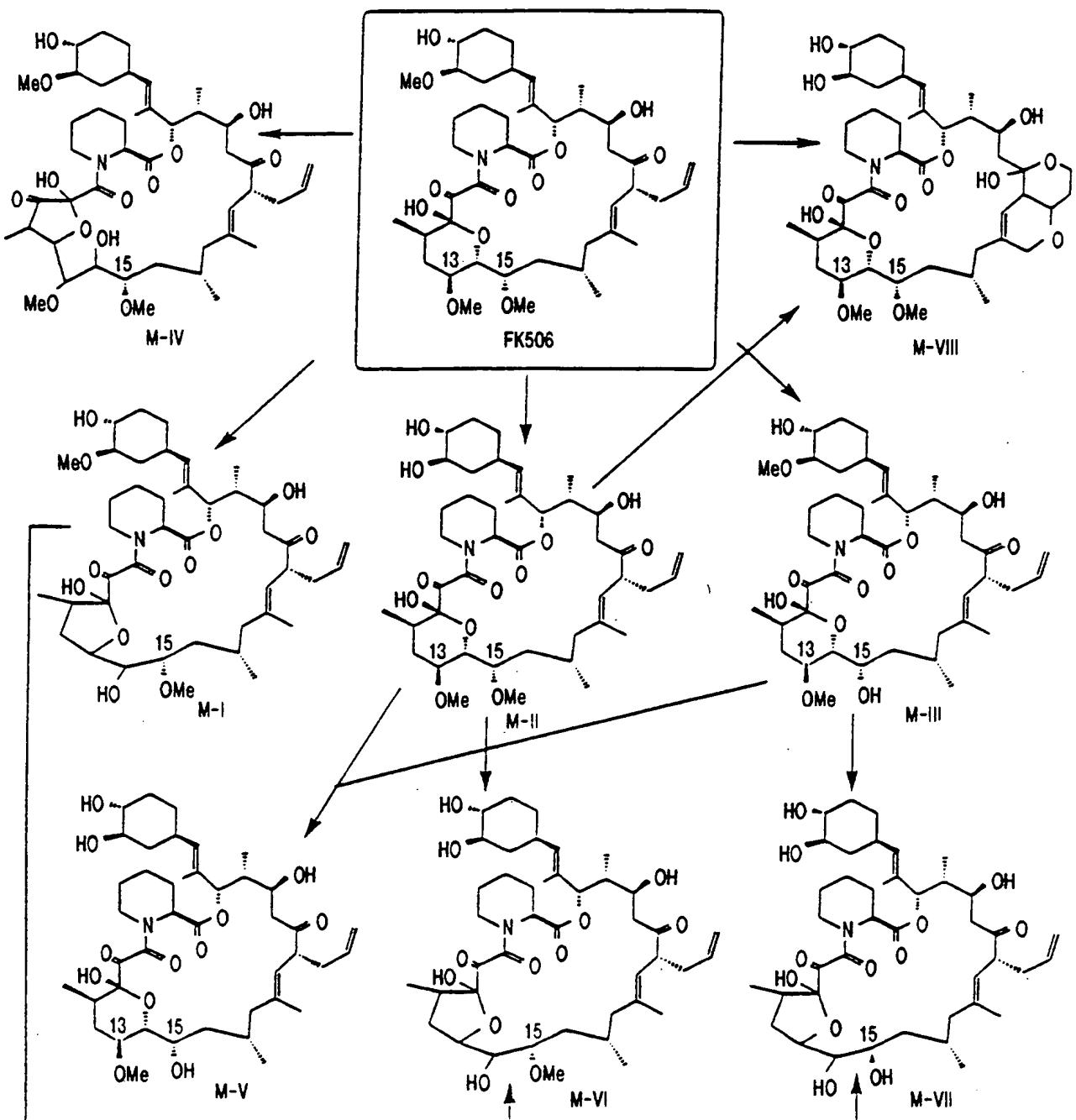


FIG. 6

7 / 9

FIG. 7A



FIG. 7B

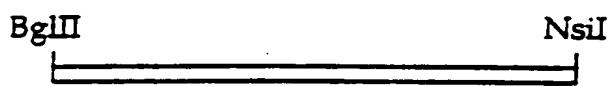


FIG. 7C

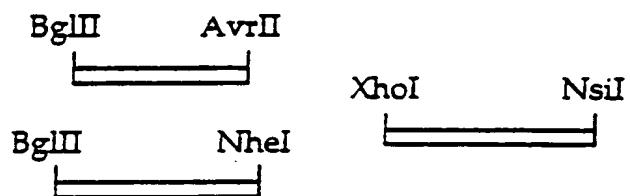


FIG. 7D

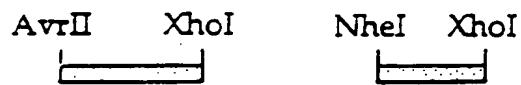
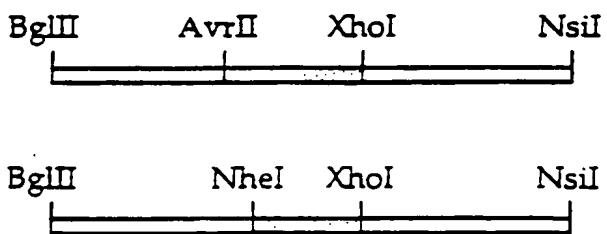


FIG. 7E



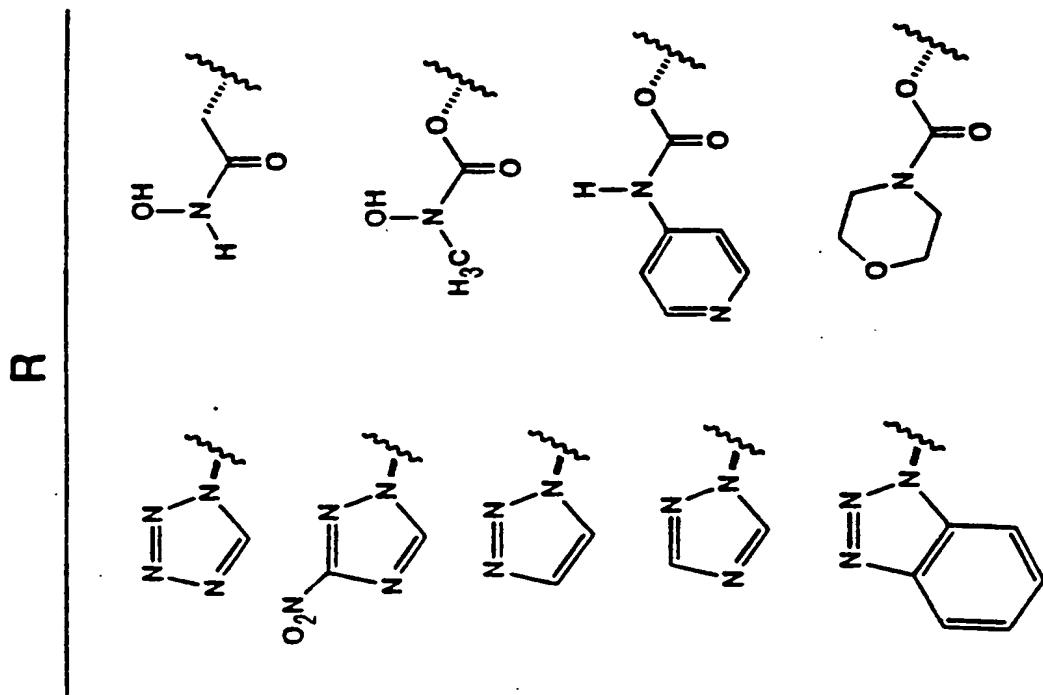
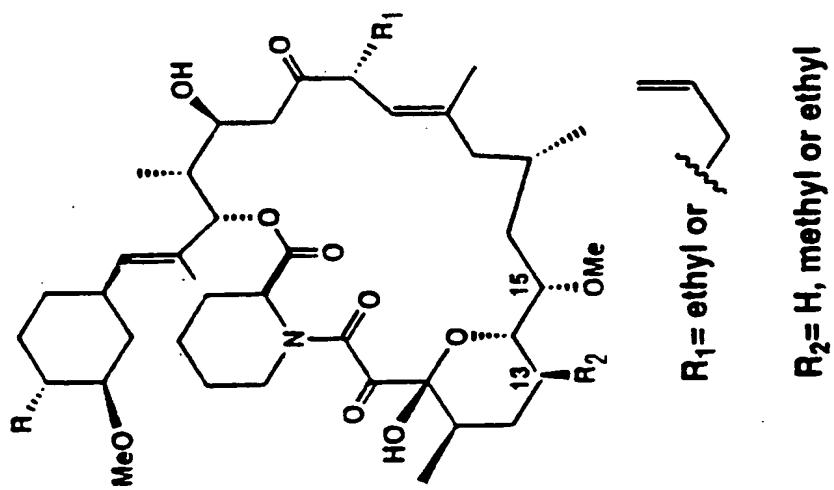


FIG. 8A



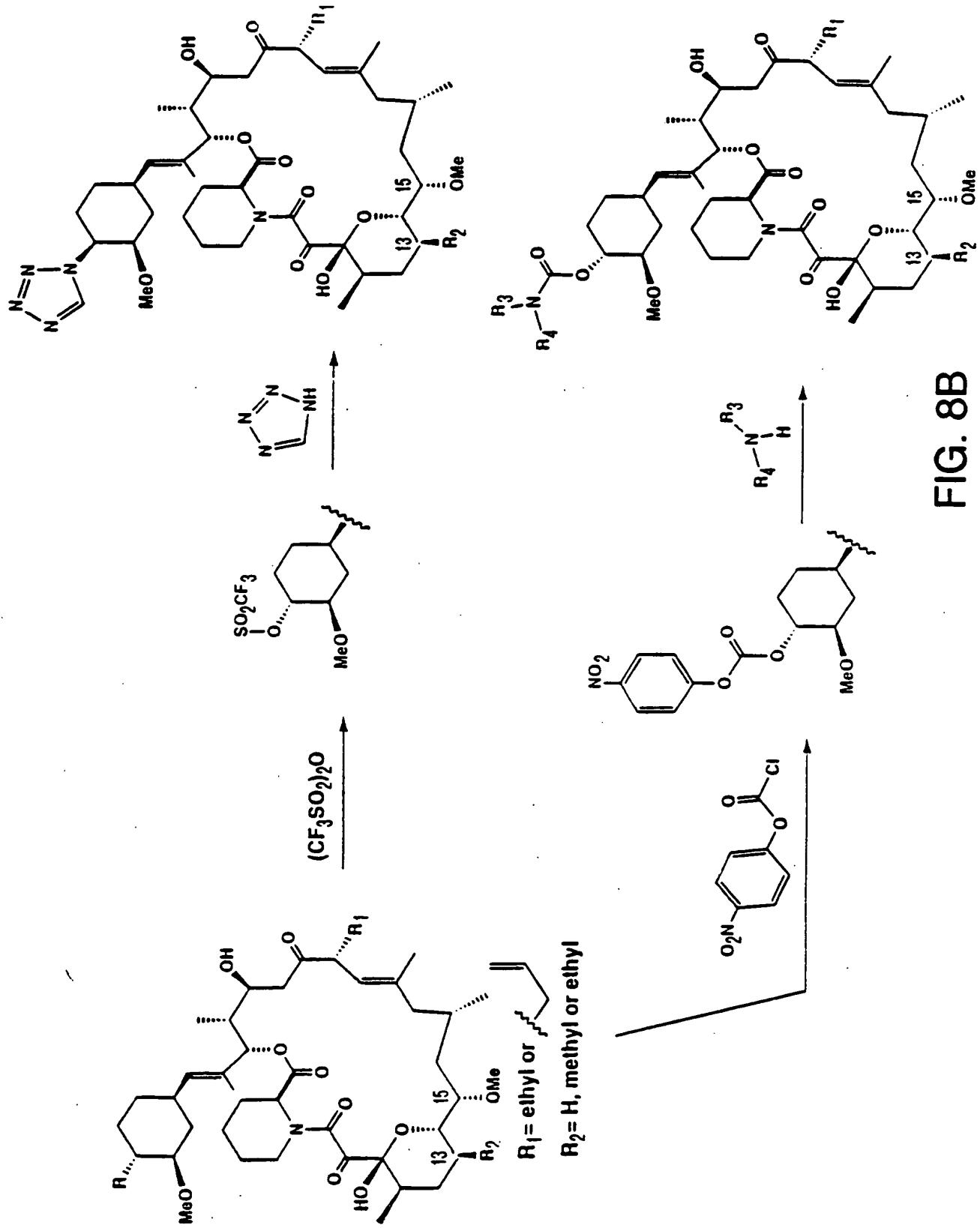


FIG. 8B

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 22, line 31-33

B. IDENTIFICATION OF DEPOSIT

Further deposits are identified on an additional sheet

Name of depositary institution

American Type Culture Collection

Address of depositary institution (including postal code and country)

10801 University Blvd
Manassas, VA 22110-2209
USA

Date of deposit

20 September 1999

Accession Number

PTA-727, PTA-728 and PTA-729

C. ADDITIONAL INDICATIONS (leave blank if not applicable)

This information is continued on an additional sheet

D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)

All designated States

E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")

For receiving Office use only

 This sheet was received with the international application

Authorized officer

For International Bureau use only

 This sheet was received by the International Bureau on:

13 JUN 00

Authorized officer

INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

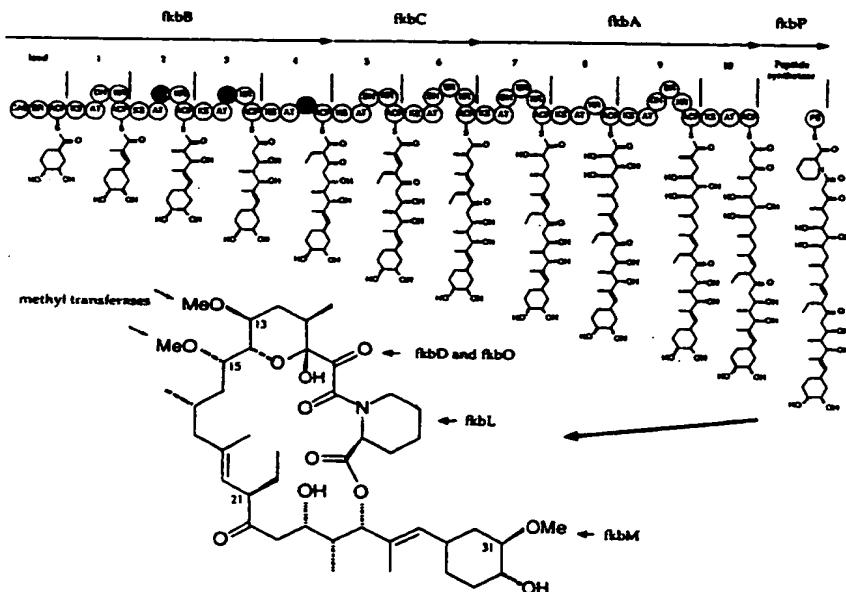
A. The indications made below relate to the microorganism referred to in the description on page <u>22</u>, line <u>31-33</u>		
B. IDENTIFICATION OF DEPOSIT		Further deposits are identified on an additional sheet <input checked="" type="checkbox"/>
<p>Name of depositary institution American Type Culture Collection</p> <p>Address of depositary institution (including postal code and country) 10801 University Blvd Manassas, VA 22110-2209 USA</p>		
Date of deposit 20 September 1999	Accession Number	PTA-726
C. ADDITIONAL INDICATIONS (Leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>		
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)		
All designated States		
E. SEPARATE FURNISHING OF INDICATIONS (Leave blank if not applicable)		
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications e.g., "Accession Number of Deposit")		
<p>For receiving Office use only</p> <p><input type="checkbox"/> This sheet was received with the international application</p> <p>Authorized officer</p>		
<p>For International Bureau use only</p> <p><input checked="" type="checkbox"/> This sheet was received by the International Bureau on: 13 JUN 00</p> <p>Authorized officer</p> <p><i>S. P.</i></p>		



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N 15/52, 15/54, 15/62, 9/10, C12P 17/18, 19/32, C07D 498/18 // (C07D 498/18, 311:00, 273:00, 211:00)		A3	(11) International Publication Number: WO 00/20601
			(43) International Publication Date: 13 April 2000 (13.04.00)
(21) International Application Number: PCT/US99/22886	(74) Agents: FAVORITO, Carolyn et al.; Morrison & Foerster LLP, 2000 Pennsylvania Avenue, N.W., Washington, DC 20006-1888 (US).		
(22) International Filing Date: 1 October 1999 (01.10.99)			
(30) Priority Data: 60/102,748 2 October 1998 (02.10.98) US 60/123,810 11 March 1999 (11.03.99) US 60/139,650 17 June 1999 (17.06.99) US	(81) Designated States: AL, AM, AU, BA, BB, BG, BR, CA, CN, CR, CU, CZ, DM, EE, GD, GE, HR, HU, IL, IS, JP, KG, KP, KR, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, SG, SI, SK, TR, TT, UA, US, UZ, VN, ZA, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).		
(71) Applicant (for all designated States except US): KOSAN BIOSCIENCES, INC. [US/US]; 3832 Bay Center Drive, Hayward, CA 94545 (US).	Published With international search report.		
(72) Inventors; and			
(75) Inventors/Applicants (for US only): REEVES, Christopher [US/US]; 4 East Altarinda Drive, Orinda, CA 94563 (US). CHU, Daniel [US/US]; 3767 Benton Street, Santa Clara, CA 95051 (US). KHOSLA, Chaitan [IN/US]; 740 Para Avenue, Palo Alto, CA 94306 (US). SANTI, Daniel [US/US]; 211 Belgrave Avenue, San Francisco, CA 94117 (US). WU, Kai [CN/US]; 900 Constitution Drive, Foster City, CA 94404 (US).	(88) Date of publication of the international search report: 26 October 2000 (26.10.00)		

(54) Title: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA CONSTRUCTS THEREFOR



(57) Abstract

Host cells comprising recombinant vectors encoding the FK-520 polyketide synthase and FK-520 modification enzymes can be used to produce the FK-520 polyketide. Recombinant DNA constructs comprising one or more FK-520 polyketide synthase domains, modules, open reading frames, and variants thereof can be used to produce recombinant polyketide synthases and a variety of different polyketides with application as pharmaceutical and veterinary products.

*(Referred to in PCT Gazette No. 35/2000, Section II)

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/22886

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/52 C12N15/54 C12N15/62 C12N9/10 C12P17/18
C12P19/32 C07D498/18 // (C07D498/18, 311:00, 273:00, 211:00)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12N C12P C07D

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, MEDLINE, STRAND, BIOSIS, EMBASE, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>MOTAMEDI H ET AL.: "The biosynthetic gene cluster for the macrolactone ring of the immunosuppressant FK506" EUR. J. BIOCHEM., vol. 256, no. 3, 15 September 1998 (1998-09-15), pages 528-534, XP000906738 abstract figures 1,2 page 532, right-hand column, line 51 -page 533, left-hand column, line 18 ---</p> <p style="text-align: center;">-/-</p>	12

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

27 July 2000

Date of mailing of the international search report

10/08/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

van de Kamp, M

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/22886

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>STASSI D L ET AL.: "Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering" PROC. NATL. ACAD. SCI. USA, vol. 95, no. 13, 23 June 1998 (1998-06-23), pages 7305-7309, XP002143632 abstract page 7308, left-hand column, line 4 -right-hand column, line 17 page 7309, left-hand column, line 24-40 ---</p>	17
X	<p>REYNOLDS K A ET AL.: "Rapamycin, FK506, and ascomycin -related compounds" DRUGS PHARM. SCI., vol. 82, 1997, pages 497-520, XP000906777 figure 3 page 502, line 7-25; figure 7 page 509-513, paragraph IV ---</p>	18
X	<p>DUMONT F J ET AL.: "The immunosuppressive and toxic effects of FK-506 are mechanistically related: pharmacology of a novel antagonist of FK-506 and rapamycin" JOURNAL OF EXPERIMENTAL MEDICINE, vol. 176, no. 3, 1 September 1992 (1992-09-01), pages 751-760, XP000906781 cited in the application abstract; figure 1 ---</p>	18
X	<p>KAWAI M ET AL.: "Structure-activity profiles of macrolactam immunosuppressant FK-506 analogues" FEBS LETTERS, vol. 316, no. 2, January 1993 (1993-01), pages 107-113, XP002143633 abstract scheme 1 table 1 ---</p>	18
X	<p>EP 0 323 042 A (FISONS PLC) 5 July 1989 (1989-07-05) example 13 ---</p>	18
X	<p>EP 0 356 399 A (SANDOZ AG ;SANDOZ AG (DE); SANDOZ AG (AT)) 28 February 1990 (1990-02-28) examples 2,3 ---</p>	18
X	<p>EP 0 463 690 A (MERCK & CO INC) 2 January 1992 (1992-01-02) example 3 ---</p>	18
	-/--	

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/22886

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No
Y	<p>MOTAMEDI H ET AL.: "Characterization of methyltransferase and hydroxylase genes involved in the biosynthesis of the immunosuppressants FK506 and FK520" <i>J. BACTERIOLOGY</i>, vol. 178, no. 17, July 1996 (1996-07), pages 5243-5248, XP002137077 abstract page 5245, left-hand column, line 1-3 figure 4</p> <p>---</p> <p>MOTAMEDI H ET AL.: "Structural organization of a multifunctional polyketide synthase involved in the biosynthesis of the macrolide immunosuppressant FK506" <i>EUR. J. BIOCHEM.</i>, vol. 244, no. 1, 15 February 1997 (1997-02-15), pages 74-80, XP000906743 abstract page 79, left-hand column, line 26-35 page 75, left-hand column, line 31 -page 76, left-hand column, line 1</p> <p>---</p>	1-11
A	<p>CHEN T S ET AL.: "Microbial transformation of immunosuppressive compounds. II. Specific desmethylation of 13-methoxy group of FK 506 and FR 9500520 by Actinomycete sp. ATCC 53828" <i>J. ANTIBIOT.</i>, vol. 45, no. 4, April 1992 (1992-04), pages 577-580, XP002143634 figure 1</p> <p>---</p>	18-20
A	<p>SHAFIEE A ET AL.: "Enzymatic synthesis and immunosuppressive activity of novel desmethylated immunomycins (ascomycins)" <i>J. ANTIBIOT.</i>, vol. 46, no. 9, September 1993 (1993-09), pages 1397-1405, XP002143635 abstract</p> <p>---</p>	18,20
A	<p>KHOSLA C: "Harnessing the biosynthetic potential of modular polyketide synthases" <i>CHEMICAL REVIEWS</i>, vol. 97, no. 7, 1997, pages 2577-2590, XP002130646 ISSN: 0009-2665</p> <p>-----</p>	

INTERNATIONAL SEARCH REPORT

List of patent family members

International Application No

PCT/US-99/22886

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP 0323042	A 05-07-1989	AT 120466 T		15-04-1995
		AU 630866 B		12-11-1992
		AU 2822889 A		05-07-1989
		CA 1339128 A		29-07-1997
		CN 1033458 A		21-06-1989
		DE 3853477 D		04-05-1995
		DE 3853477 T		09-11-1995
		DK 387889 A		08-08-1989
		EP 0346427 A		20-12-1989
		ES 2071681 T		01-07-1995
		FI 90550 B		15-11-1993
		FI 930597 A		11-02-1993
		WO 8905304 A		15-06-1989
		IE 66163 B		13-12-1995
		IL 88629 A		12-04-1994
		JP 2502463 T		09-08-1990
		JP 2799208 B		17-09-1998
		NO 893166 A		04-08-1989
		NZ 227251 A		26-02-1990
		PT 89203 A, B		29-12-1989
		US 5376663 A		27-12-1994
		ZA 8809136 A		30-08-1989
EP 0356399	A 28-02-1990	AU 629563 B		08-10-1992
		AU 4024689 A		01-03-1990
		DK 418789 A		27-02-1990
		JP 2167287 A		27-06-1990
		NZ 230418 A		25-10-1991
		US 5011844 A		30-04-1991
		ZA 8906524 A		24-04-1991
EP 0463690	A 02-01-1992	CA 2044846 A		26-12-1991
		DE 69117221 D		28-03-1996
		DE 69117221 T		12-09-1996
		JP 4230288 A		19-08-1992
		US 5190950 A		02-03-1993
		US 5342935 A		30-08-1994

OIP/E/JCWS
MAY 13 2002
RECEIVED